Author's Response

to the comments of Referees and the Editor of NHESS regarding the discussion paper

Kämäräinen, M., Hyvärinen, O., Jylhä, K., Vajda, A., Neiglick, S., Nuottokari, J., and Gregow, H.: A method to estimate freezing rain climatology from ERA-Interim reanalysis over Europe, Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-225, in review, 2016.

In Helsinki, 4 January 2016
Matti Kämäräinen

Table of Contents

1. A point-by-point response to the reviews

Please note that the comments given here by the author are partly different than the comments that were sent to the 'Public discussion' of the paper.

Referee #1, Dimitar Nikolov

National Institute of Meteorology and Hydrology 1784 Sofia, Bulgaria

*We would like to thank Dr. Nikolov for the in-depth review and detailed comments. His comments are in **bold** and our replies to the comments are in normal font. We included here and replied only the critical comments.*

…

Specific comments

I would recommend the authors replace the references Rauber et al., 2000; Carrière et al., 2000 for the warm rain process (page 2) with (or add) the following two:

- **Bocchieri, J., 1980: The objective use of upper air soundings to specify precipitation type. Mon. Wea. Rev., 108, 596–603.**
- **Huffman, G. J., and G. A. Norman Jr., 1988: The supercooled warm rain process and the specification of freezing precipitation. Mon. Wea. Rev., 116, 2172–2182.**

We kept Rauber et al. (2000) and replaced Carrière et al. (2000) with Bocchieri (1980), and Huffman and Norman (1988).

I disagree with the decision of the authors to exclude most of the stations from eastern Europe. It seems that they have erroneously interpreted the explanations of Bezrukova et al. (2006) for the different definitions of FZRA events in these countries. Indeed, sometimes the icing due to supercooled clouds or fogs may deposit as glaze (wet growth process) and then the symbol for glaze is written

down, but such a case will never be reported as freezing rain or freezing drizzle in the WMO weather codes. This ambiguity concerns mostly the local meteorological archives where additional control is needed to distinguish between both events. The weather codes 24, 56, 57, 66 and 67 are not affected at all. By this reason the authors (of Bezrukova et al., 2006) have decided to restrict only to the WMO codes.

Indeed, it seems to be likely that we misinterpreted the methodology of Bezrukova et al. (2006). In December 2016 we contacted Dr. Bezrukova via email and she confirmed that in Russia, the SYNOP station observations are performed following the WMO standards.

Based on our results, though, it is clear that when all stations from the domain (i.e. including eastern stations) are used in calibration of the FMICLIM algorithm, we get a prominent overestimation of algorithm-based total number of FZRA events in eastern parts and underestimation in other parts of Europe for some reason.

Because we could not identify whether the reason to the differences in predicted and observed values originates from the reanalysis or from observations, and following the suggestions of Drs. Nikolov and Bezrukova, we extended the station set to include also the eastern stations. We recalibrated the algorithm with all relevant stations and modified the text, tables, and figures accordingly. The most important effect was found in the mean values, so that recalibrated algorithm produces less freezing rain than the previous version. Additionally the validation results deteriorated: some metrics, e.g. the correlation between the spatially averaged observed and predicted annual mean numbers is lower after calibration than before it.

The authors of the manuscript have also filtered the data outside the interval –30oC +10oC, which seems to be too wide. Most often FZRA occur in the interval –10oC 0oC, so an appropriate interval for filtering, in my opinion, would be –15oC + 5oC. This would prevent to a certain extent from misclassification of ice pellets as FZRA or FZDR.

We recalibrated the algorithm using this new stricter filter, which enhanced the validation results slightly. The enhancement was not as significant as the deterioration of the results following the inclusion of the eastern stations.

The finding that the altitude does not contribute to the explain variance is somehow surprising for me. One would expect that the number of FZRA and their

duration would decrease with the altitude because of the decreasing of the depth of the near-surface cold layer and FZRA aloft should be even more rare event than the FZRA at the ground. However, mountain ranges mostly caused cold air damming which is difficult to be recognized in data sets with coarse resolution.

The text was clarified so that the correlation of distance and elevation is more clear. The highest elevations, where freezing rain is very rare, were masked out from the maps and analyses, because the algorithm can not work there due to lack of pressure levels.

The vertical resolution of the FMINWP seems to be not very appropriate for detailed representation of the vertical profiles of the relative humidity and the air temperature, which would affect the correct estimation of the near-surface cold layer and the melting layer above. It can be seen that an increasing of the resolution is foreseen as future work and this would be very helpful.

Using a coarse vertical resolution, despite the drawbacks, was partly selected because output data from climate models, which we are going to analyse next, is commonly available in a rather coarse vertical resolution as well, and one purpose of this paper was to show that some kind of results can be achieved also by that way.

Some clarifications were added to the Introduction and Conclusions parts.

The minimum acceptable cold layer depth has been significantly increased by the calibration procedure – from 130 meters up to 400 meters. This seems very reasonable because of the large size of the investigated area and variable weather conditions. For example Bernstein reported values of the near-surface cold layer in USA between 100 and 1400 meters, the minimums being between 100 and 300 meters.

- **Bernstein, B., 2000: Regional and local influences on freezing drizzle, freezing rain, and ice pellet events. Wea. Forecasting, 15, 485–508.**

We completed the text based on this comment, and added the new provided reference.

…

Very interesting results are presented in the paragraph 3.3 Climatology of freezing rain in Europe. However, the finding for a maximum in the annual number of events over the Carpathian mountain sounds surprisingly for me. It would be useful if the altitude of these regions is given.

The highest elevations, where freezing rain is very rare, were masked out from the maps, because the algorithm can not work there due to lack of pressure levels. Also, elevation information was included in the Fig. 1.

Technical comments

I have encountered only two small misprints – on page 9, third row – the FMICLIM is written wrongly and on page 11, third row is written "The Carpathian...".

This was corrected.

Conclusions

…

The size, quality and readability of each figure is adequate to the type and quantity of data presented, except for figure 5 which could be a little bit larger.

The figure was enlarged slightly.

The authors give proper credit to previous and related work with a small oversight of two references for the warm rain process.

The references were adjusted based on this comment.

…

The authors should only take into account that they could use for future investigation the stations in eastern Europe with no restrictions, as far as they utilize the international weather WMO codes.

The method was recalibrated, the results adjusted, and figures replotted using all relevant stations.

…

Referee #2, anonymous

*We would like to thank the Referee #2 for the in-depth review and detailed comments. His/Her comments are in **bold** and our replies to the comments are in normal font. We included here and replied only the critical comments.*

…

Specific Comments

Page 1, Line 17: There are many instances of short but 'heavy' freezing rain.

'Short-lived' was removed from the text.

Page 2, Line 12: A recent climatology over parts of northern Eurasia has been completed:

- **Groisman, P.Ya., O. N. Bulygina, X. Yin, R. S. Vose, S. K. Gulev, I. Hanssen-Bauer, and E. Førland 2016: Recent changes in the frequency of freezing precipitation in North America and Northern Eurasia. Environ. Res. Lett. 11, 045007.**

This work was included as a reference.

Page 3, Line 27: 3-hourly reports are probably insufficient. Most freezing rain events occur at shorter time scales. Was any attempt made to at least 'estimate' how many events were uncounted by using hourly information as well?

No such attempt was made since we did not have hourly observations available. Without good observations, too many assumptions are required for the estimation of the distribution of the short-term events. This would be an interesting exercise to do, but it is out of the scope of the present article, and indeed requires 1-hourly (or maybe even denser) observations.

Page 3, Line 27: I may have missed this but how did you treat combinations of precipitation types? It is common for freezing rain to occur with ice pellets for example.

According to WMO standards, only the weather with the largest code number is reported in the SYNOP present weather code. Therefore, if e.g. ice pellets (79) occur together with freezing rain (66, 67), only ice pellets are reported, and no information of the simultaneous FZRA is recorded at all. We calibrated the FMICLIM method to identify reported FZRA, and for that reason selected SYNOP codes 24 (freezing rain within past hour but not at observation time), 66 (light freezing rain) and 67 (moderate to heavy freezing rain) as target classes.

Page 4, Line 3: A threshold of 80% is quite low. Why didn't you show the fraction of missing data during the cold season?

The fraction of missing data during the cold season is now presented in the text to help readers to estimate the number of missing FZRA observations in the total numbers of FZRA events.

Page 4, Line 15: What fraction of observations was beyond these thresholds? Were any of the high valued temperatures associated with very low relative humidities that would lead to much lower wet bulb temperatures?

The algorithm was recalibrated using the stricter temperature intervals, which enhanced the calibration results slightly.

The combination of high temperatures and very low relative humidity is quite rare in the observational data:

- 21% of the freezing rain happened with above-zero temperatures
- 5% happened with T2m > 0C AND RH2m < 90%
- 1% happened with T2m > 0C AND RH2m < 80%
- 0.1% happened with T2m > 0C AND RH2m < 60%

Of course, using wet bulb temperatures instead of air temperatures would most probably enhance the results. This was added to the 'Future work' section in the text.

Page 5, Line 7: This scale is very large for freezing rain. It is quite common for these regions to be less.

We would prefer to use spatially and temporally denser reanalyses. As said in the text, this is considered to be one part of the future work.

Page 6, Line 3: Ice-initiated precipitation is not initially generated in the inversion aloft. What is the implication of assuming it is?

The implication is that the algorithm does not identify those FZRA events. Those non-identified events are compensated in the calibration by adjustments of the temperature thresholds in different layers. These adjustments then lead to generation of erroneous FZRA events to some other time steps (i.e. false alarms), and to deterioration of the validation results.

The separation of moist (precipitation generation) layer and melting layer are mentioned in the 'Future work' section.

Page 11, Line 31: This paragraph is poorly worded and hard to follow.

The paragraph was restructured.

Page 12, Line 5: There are standard observing practices to identify freezing rain. Why is this so hard to do?

As said in the text, simply "because the phenomenon can be easily confused with ordinary, non-freezing rain": For an observer, freezing rain looks the same by eye as non-freezing rain, and accumulation of ice might not be visible (1) in short-term events or (2) when snow covers objects on the ground. Besides, the observed 2m temperature, which is most commonly used by the observers to distinguish FZRA from non-freezing rain at stations, might not represent the temperature of the thick (maybe several hundreds meters) near-surface cold layer well enough in all cases, and knowing if the rain actually is supercooled or not is not so straightforward.

Page 12, Line 8: You are associating 'minor' with short duration. On what basis? There can be severe impacts with durations smaller than 6 h and precipitation rates can be high as well..

References to 'minor FZRA' in the text were removed.

Page 12, Line 11: Are the errors 'random'?

They are 'random' in the sense that we try to predict station level variability using grid cell level information. Also they are 'random' in the sense that we assume the observational errors to be random so that equal amounts of (1) false identifications of FZRA events and (2) false rejections of FZRA events happen.

The paragraph was completed to be more clear.

Page 12 Line 25: Given the enormous smoothing at 70 km, maybe the authors should only consider analyses over 'flat regions'?

The highest elevations were excluded from the analysis and masked out from the maps.

Page 12, Line 33: "Occasional misclassification"? How often did this occur?

It happens sometimes, as shown in the Section 3.2.1 (SYNOP weather code classification) and in Fig. 6.

Page 13, Line 3: To me, this section is too long and wordy. This is a long shopping list. What are the most important and feasible next steps? From my perspective, some of these should be done within this article.

The list was ordered and modified based on the expected importance of the future steps.

As well, a recent article (Liu et al., 2016) pointed out that precipitation at the surface (including freezing rain) is calculated directly from the model's microphysical package without needing the approach used here. Isn't that the best way forward?

- **Liu et al., 2016: Continental scale convection permitting modeling of the current and future climate of North America. Climate Change, DOI 10.1007/s00382-016-3327-9**

This is of course the optimal solution. However, the precipitation type is not included in the output variables of ERA-Interim (or in the publicly available climate model output datasets which we are going to analyse next).

Page 14, Line 12: I do not think that it is 'sophisticated enough...'. Melting rates of particles aloft, for example, depend on the features of the particles themselves as well as temperature and moisture conditions.

'Sophisticated enough' was removed.

Page 14, Line 15: Why did you not examine sounding information taken during freezing rain events? You could then more quantitatively assess how well the approach is handling particular instances. The lack of such validation is a major drawback in this article.

Examining sounding information was given a high priority in the list of the future steps (Sec. 4.2).

Page 14, around Line 25: Why not compare against previous studies on the climatological features?

The most relevant studies in this context, i.e. Bezrukova et al., 2006, Carriere et al., 2000, and Groisman et al., 2016, can not be compared to our results because of different methodology (Bezrukova et al., 2006 and Groisman et al., 2016) or too a short period used (Carriere et al., 2000).

Page 14, Line 30: Given the limitations of the dataset as you have mentioned, how confident are you that you can 'reliably' address such questions?

In that sentence, we (implicitly) refer to Fig. 3, and there we can be quite confident because our results are backed by the observations. There are other results that we are less confident and which are a motivation for further development of the FZRA detection methodology.

Page 14, Line 31: Clarify what is meant by 'station scale analysis'.

'Station scale' was replaced with 'station level' in all occurrences (i.e. in the Conclusions and the Abstract sections).

Station level analysis means comparisons of raw modelled and observed time series of FZRA in all individual stations. No aggregation of data temporally or spatially is applied prior to analysis.

Technical Corrections

Page 2, Lines 5 and 7: The word ‘where’ is not correct in referring to an event ‘in time’. This error was done in other places as well.

This was corrected.

Page 2, Line 22: Another incorrect use of ‘where’.

This was corrected.

Page 13, Line 28: ‘criteria’

This was corrected.

2. A list of all relevant changes made in the manuscript

1. The calibration was recalculated using all relevant weather stations, i.e. the eastern stations which were previously excluded, we included in the calibration.
2. A new figure was added to clarify the east-west bias results in the station network.
3. Figures were replotted using the newly calibrated algorithm, and the highest elevations were masked out in the maps.
4. Numbers in the text and tables were corrected after recalibration.
5. The chapter ‘Number of freezing rain events at weather stations’ was partly restructured and the detailed explanations of changes in statistics were removed, because the calibration did not change them anymore after the adjustments in calibration station set.
6. The chapter ‘Future work’ was partly restructured.

3. A marked-up manuscript version

Please note that the figures in this marked-up version are partly old, and partly wrongly numbered. Additionally the ‘latexdiff’ software adds some extra lines in the beginning of the document for some reason.

The ~~monthly climatology~~ highest elevations were excluded (gray color) because of ~~FZRA~~ larger uncertainties in ~~1979–2014 according to the FMI$_{CLIM}$ algorithm, applied to the ERA-Interim. The average annual number of 6-hourly~~ FZRA ~~cases is shown~~detection. figure.9

A method to estimate freezing rain climatology from ERA-Interim reanalysis over Europe

Matti Kämäräinen[1], Otto Hyvärinen[1], Kirsti Jylhä[1], Andrea Vajda[1], Simo Neiglick[1], Jaakko Nuottokari[1], and Hilppa Gregow[1]

[1]Finnish Meteorological Institute, Erik Palménin aukio 1, P.O.Box 503, FI-00101 Helsinki, FINLAND

Correspondence to: Matti Kämäräinen (matti.kamarainen@fmi.fi)

Abstract.

A method for estimating the occurrence of freezing rain (FZRA) in gridded atmospheric datasets was evaluated, calibrated against SYNOP weather station observations, and applied to the ERA-Interim reanalysis for climatological studies of the phenomenon. The algorithm, originally developed at the Finnish Meteorological Institute for detecting the precipitation type in numerical weather prediction~~at the Finnish Meteorological Institute~~, uses vertical profiles of relative humidity and temperature as input. Reanalysis data in 6-hourly time resolution was analyzed over Europe for the period 1979–2014. Mean annual and monthly numbers of FZRA events, as well as probabilities of duration and spatial extent of events, were then derived. The algorithm was able to reproduce accurately the observed, spatially averaged interannual variability of FZRA (correlation ~~= 0.93~~0.90) during the 36-year period, but at station level rather low validation and cross-validation statistics were achieved (mean correlation ~~= 0.44~~0.38). Coarse grid resolution of the reanalysis, and misclassifications to other freezing phenomena in SYNOP observations, such as ice pellets and freezing drizzle, contribute to the low validation results at station ~~scale~~level. Although the derived gridded climatology is preliminary, it may be useful, for example, in safety assessments of critical infrastructure.

1 Introduction

Freezing rain (FZRA) is liquid, supercooled precipitation which freezes on coming into contact with solid objects, forming a coating of ice (World Meteorological Organisation, 2010, 2011). It is a relatively rare but high-impact wintertime weather phenomenon, and in Europe it affects mainly central, eastern and northern parts of the continent. Although major events resulting in heavy ice accretion are not as common as lighter ~~, short-lived~~ cases, the direct damages they cause to critical infrastructure (transportation, communication and energy), and forestry are substantial. For example, the ice coating formed on trees and powerlines causes them to fail, leading to severe power outages, transportation disruption, delays in emergency responses, and severe economic losses (Call, 2010; Lambert and Hansen, 2011). Lighter freezing rain events are also harmful

because of their indirect effects, the most important being the reduced friction on road surfaces that results in increased rates of accidents, injuries, and difficulties in transportation (Degelia et al., 2015).

During recent years some major events have been experienced across Europe. On 31 January – 3 February 2014 a prolonged, heavy FZRA and blizzard event hit the Alpine region, Hungary and the Balkan Peninsula. In Slovenia, during these four days, 40–300 mm precipitation fell over and resulted in 10 cm accumulation of ice (Markosek, 2015). Severe damage was caused to critical infrastructure, e.g. 30 km of power lines were completely destroyed and 174 km were inoperative, while railways and road traffic were heavily disrupted or closed for several days. Over 0.5 million hectares of forest were damaged, and the total cost resulted in around 400 million euros (Vajda, 2015). In Croatia, over 80% of the population were left without electricity (Editorial, 2014). Other severe cases include the Moscow FZRA event on 25–26 December 2010, where flights were cancelled at the Domodedovo airport and power supplies to trains, trams, and busses were destroyed; and the 13 December 2012 case, where the British Isles and France suffered from a mix of freezing rain and snow.

Despite the severe impacts of FZRA events, only a few publications concerning the European climatology of this phenomenon exist currently. Carrière et al. (2000) provided a climatology of freezing precipitation (including FZRA, freezing drizzle and ice pellets) based on SYNOP weather station reports from three winters. Bezrukova et al. (2006) and Groisman et al. (2016) presente climatological information of FZRA over ~~the former USSR, thus including~~ Russia and eastern Europe. Most of the other studies focus on the occurrence of in-cloud or near-surface icing (e.g., Bernstein et al., 2009; Bernstein and Le Bot, 2009; Le Bot, 2004; Le Bot and Lassegues, 2004). More comprehensive studies have been undertaken on ice storm and FZRA climatologies over North America, including studies for various regions (Cortinas, 2000; Changnon, 2003; Cortinas et al., 2004), impacts of ice storms on different sectors (Proulx and Greene, 2001), and changes in FZRA climatology (Cheng et al., 2007, 2011; Lambert and Hansen, 2011; Klima and Morgan, 2015) as well as various wintertime precipitation types (Stewart et al., 2015), and impacts and details of FZRA storms and related synoptics (Hosek et al., 2011; Call, 2010; Roberts et al., 2008).

The two formation mechanisms of FZRA are rather well known. In the majority of cases a near-surface freezing layer with an accompanying melting layer above makes the hydrometeors – formed above these layers – to be in liquid, supercooled phase when they hit the ground and freeze on contact with objects at the surface (World Meteorological Organisation, 2010). Other FZRA cases occur without the cold layer – melting layer structure, as a result of the warm rain process ~~(Rauber et al., 2000; Carrière et al., 2000)~~(Bocchieri, 1980; Huffman and Norman, 1988; Rauber et al., 2000), where collision and coalescence of the small droplets ensure the liquid form. The latter mechanism is usually associated with drizzle or freezing drizzle, but in some cases it may lead to formation of FZRA.

Several approaches have been developed for identifying wintertime precipitation types (e.g., snow, ice pellets, freezing rain) in numerical weather prediction (NWP) models, most of them for North America and many of them reviewed by Cortinas et al. (2002). With varying complexity, all of them are based on the vertical temperature profile which is used to predict the state of the hydrometeors in the atmosphere and on the surface level in particular. Usually the vertical humidity profile is used as well. For example, Ramer (1993) developed an empirical method which explicitly resolves the melting and freezing of the descending hydrometeors. This method has been widely used in NWP and in related studies (e.g. Reeves et al., 2014), showing a good skill among the other precipitation typing algorithms. A slightly more indirect, and perhaps simpler approach

was presented by Bourgouin (2000), who estimated the phase of precipitation based on areas between the temperature profile and the 0°C isotherm on a tephigram.

Three-dimensional gridded meteorological datasets at daily or sub-daily temporal resolution, such as output from numerical climate models or reanalysis models, are commonly used in climate studies to account for gaps in time series of weather station data, and to fill the sparsely covered areas, like seas and the above-surface atmosphere. However, a variety of issues complicate their use for estimating precipitation types. One important uncertainty arises from the fact that it is not straightforward to compare point-like weather station observations (representing local climate) and grid cells (representing climate of a larger area). As shown e.g. by Stewart et al. (2015), Reeves et al. (2014) and Ryzhkov et al. (2014), even small details in the vertical distribution of temperature can affect the surface precipitation type. However, because a gridded dataset typically has a rather coarse ~~spatial resolution and a smoothed orography~~horizontal, vertical, and temporal resolutions, its vertical temperature structure may differ locally from the reality. Similarly, very minor modelling uncertainty, or natural uncertainty related to subgrid-scale processes, might cause the predicted temperatures of the freezing or melting layers to be slightly off from the values that would lead to FZRA.

In this study we introduce a freezing rain detection algorithm, originally developed and operationally used in the numerical weather prediction at the Finnish Meteorological Institute (denoted by FMI_{NWP}), and here implemented for climatological applications. The algorithm was applied to the ERA-Interim reanalysis data (Dee et al., 2011) to test the applicability of a coarse-resolution method in derivation of proxies of FZRA, and to provide a climatology of FZRA in Europe for the period 1979–2014. First, the SYNOP weather station and reanalysis datasets used in calibration of the algorithm are introduced, along with an optimization-based calibration procedure; second, the calibration results are validated using multiple approaches; and lastly the climatology is produced and analysed shortly. In the analysis the following statistics are focused on: the total number of 6-hourly FZRA cases at each station or grid cell during the 36-year study period, and the average frequency of these cases per location in the whole study domain and in three subregions as a function of time. In addition, duration and spatial extent of the FZRA events are studied.

2 Materials and methods

The FMI_{NWP} algorithm uses threshold values of the air temperature and humidity in the near-surface freezing layer and in the above melting layer to distinguish the FZRA events from non-events. In order to estimate the FZRA climatology in Europe based on reanalysis data, these threshold values needed to be reconsidered for two reasons. First, some of the original threshold values were subjectively selected by meteorologists in Finland, which involves uncertainty related to subjective decisions. Second, the values are likely to be somewhat sensitive to potential biases in temperature and humidity, and these biases may be different in NWP data and in reanalysis data. Consequently, a calibration procedure was developed that employed SYNOP weather station observations to redefine the threshold values of the parameters, as discussed below. The calibrated version of the FMI_{NWP} algorithm, used in the climatological analysis of FZRA, is denoted here by FMI_{CLIM} ~~algorithm~~.

2.1 SYNOP weather station data

For validation and algorithm-calibration purposes, the observed occurrence of FZRA events was derived from 3-hourly SYNOP weather station recordings. Data from ~~4000~~ 4600 manually operated stations were collected from the Meteorological Archival and Retrieval System (MARS) of the European Centre for Medium-Range Weather Forecasts (ECMWF). Automated stations were not accepted due to their reduced ability to distinguish different types of precipitation~~(Sheppard and Joe, 2000)~~, especially freezing rain and freezing drizzle (Marjin, 2007). The present weather part of a SYNOP observation (World Meteorological Organisation, 2011) consists of one hundred codes describing the most important weather at the time of observation and one hour before it. In this study the WMO codes directly referring to FZRA were selected to represent the phenomenon: 24 (freezing rain within past hour but not at observation time), 66 (light freezing rain) and 67 (moderate to heavy freezing rain).

To be included in this study, the stations were required to contain a valid present weather code in >80% of the 3-hourly time steps during the period 1979–2014. If stations with shorter or less regular records had not been excluded, the observed number of FZRA events per station might have been distorted. Besides, regularly working and maintained stations with high-frequency observations are assumed to be more reliable. Altogether 525 stations out of ~~4000~~4600, presented in Fig. 1, passed these first conditions. During the cold season months September – May the proportion of missing data was 10% on average.

For calibration and validation purposes, and for further analyses of FZRA in station locations, the data was filtered further:

– ~~In many countries of eastern Europe, the FZRA has been interpreted differently compared to other European countries, so that FZRA is reported only in the presence of simultaneous glaze ice (Bezrukova et al., 2006). This leads to underestimation of the observed number of cases, and for this reason most of these stations were excluded.~~

– ~~The stations~~ Stations located above 2000 m above sea level ~~and those having less than 10 FZRA observations were excluded ,~~ (8 stations) were excluded as the algorithm does not have enough pressure levels for high elevations~~, and~~ .

– Stations having less than 10 FZRA observations (224 stations) were excluded because reliable observations might be difficult for observers with limited experience of the phenomenon due to its rarity.

– To exclude grossly erroneous recordings, the FZRA observations with the surface temperatures below ~~-30~~-15°C or above +~~10~~5°C were rejected.

Applying the above mentioned restrictions excluded ~~325~~ 224 stations (circles in Fig. 1) and thus reduced the total number of stations to ~~200.~~ 293. These remaining stations ~~, located predominantly in northern and central Europe, are~~ were used in the calibration and validation process of this study. Finally, the time steps 00Z, 06Z, 12Z and 18Z were picked from the 3-hourly SYNOP observations to allow direct comparisons of the present weather codes with predictions of the FZRA events that were derived with the FMI_{CLIM} algorithm using the 6-hourly ERA-Interim reanalysis data. The comparisons were conducted both for the individual stations and for spatially clustered stations.

In order to divide the stations into clusters, variables which best explain the spatial differences in the total number of observed FZRA cases in ~~1979–2014~~ 1979–2014 were seeked. Therefore, different linear models were fitted. Strictly speaking, the number of cases are count data (non-negative integer values from counting) and should be modelled using the nonlinear Poisson regression, but this did not change our conclusions and linear models are somewhat easier to understand. Finally, the

best variables were used to classify stations into subgroups, and further analyses were performed for these groups separately. The variables studied were the distance to the nearest coastline (NASA, 2009), station elevation, and longitude. ~~Modelling~~ When modelling the number of cases with only one variable, the distance ~~from the coastline was the best~~ to the coastline and elevation were the best ones in explaining the variance (~~the adjusted $R^2 = 0.13$~~e.g. the adjusted R^2 for distance was 0.06, $p << .001$). ~~Adding another variable or~~ Because elevation and distance are rather strongly correlated, using both or their interaction term did not improve the results substantially, ~~while a simple stepwise model selection using Akaike information criterion (Venables and Ripley, 2002) selected also the station elevation, the adjusted R^2~~ that is, it did not change ~~. To keep the analysis simple~~the adjusted R^2. For this reason, only the distance to the nearest coastline was ~~thus~~ used for the classification. The stations were ~~then~~ grouped to "coastal" (0–140 km), "semi-coastal" (140–330 km), and "continental" (>330 km). Boundaries of classes~~were selected simply~~ , shown in Fig. 1, were selected so that each group contained equal number ~~(67)~~ of validation stations. ~~Boundaries of classes are shown in Fig. 1.~~

2.2 ERA-Interim reanalysis data

Relative humidity and temperature from 925, 850, 700 hPa and 2-meter levels, surface pressure, and precipitation of the ERA-Interim reanalysis dataset (Dee et al., 2011) were used as a predictor data. ~~First, the data was bilinearly interpolated to station locations for~~ The variables were derived from the 6-hourly analysis part of ERA-Interim except that for precipitation the 6-hourly forecasted part was used and the 12-hourly precipitation sums in were transformed into 6-hourly sums. For calibrating and validating the FZRA detection algorithm, ~~and after calibration, the~~ the data was bilinearly interpolated to station locations. The climatology of FZRA was derived using the gridded data in the original $0.7° \times 0.7°$ resolution. ~~Only the 6-hourly analysis part of the data was used except for the precipitation, where the 6-hourly forecasted part was used. The 12-hourly precipitation sums were transformed into 6-hourly sums.~~

When the predicted and observed FZRA events were compared at station level, time steps without SYNOP observations of present weather codes were excluded also from the interpolated reanalysis data, which ensured the comparability. Even though excluding those time steps leads to an underestimation of the total number of FZRA cases, the effect is ~~presumably minor, as the stations were selected so that their time series were required to be at least 80% complete~~minor, because the proportion of missing data was only 10% on average during winter months.

2.3 FMI$_{NWP}$ algorithm

The FZRA identification part of the precipitation typing algorithm (FMI$_{NWP}$), used at the weather service of the Finnish Meteorological Institute for numerical weather predictions, was adopted in this study. The algorithm uses temperature and relative humidity from four pressure levels (surface, 925, 850 and 700 hPa), and surface air pressure. Surface pressure is used to avoid analysing below-surface data in mountainous regions. The FMI$_{NWP}$ algorithm is originally used to predict locations where FZRA is conceivable, and as such it does not take the modelled precipitation intensity (Pr) into account. In our analyses the precipitation intensity was included to identify the actual FZRA cases. For FMI$_{NWP}$ the Pr value ~~0.04~~ 0.15 mm 6h^{-1} was

selected so that the identified total number of FZRA cases, calculated as a sum over the whole study domain and all the years, corresponds the observed total of ~~7900~~ 11000 cases in the ~~200~~ 293 stations.

A pseudocode representation of the algorithm is shown in Fig. 2. First, the preconditions for FZRA are checked: (1) the near-surface air temperature T2m = T_{cold} has to be lower than its predefined threshold value T_{cold}^{thr}, and (2) the maximum temperature of the above-surface layers T_{max} and (3) the surface precipitation rate Pr need to be higher than their threshold values (T_{melt}^{thr} and Pr^{thr}, respectively). In the next step the upper level of the near-surface cold layer p_{cold} is defined by selecting the pressure level closest to the ground surface, while additionally taking into account the minimum acceptable cold layer depth h_{cold}^{thr} (in hPa) which ensures that the falling raindrops are properly supercooled. Finally the existence of a moist and warm melting layer is checked by investigating the layers above the cold layer. If temperature and humidity RH in at least one of those layers are above their predefined thresholds T_{melt}^{thr} and RH_{melt}^{thr}, FZRA is predicted.

Compared to other precipitation detection algorithms, such as those presented by Ramer (1993) and Bourgouin (2000), the FMI$_{NWP}$ is presumably faster to implement and run, which makes it ideal for analysing large climatological datasets. Of these two algorithms, FMI$_{NWP}$ resembles Bourgouin (2000) more, mainly because the depth and temperature of the near-surface cold layer ($h_{cold} = p_{surf} - p_{cold}$ and T_{cold}, respectively) together describe the energy required to supercool the raindrops. FMI$_{NWP}$ assumes that melting layer and the layer where precipitation is generated are the same, even though in reality they can be separated so that precipitation is formed above the melting layer.

2.4 Calibration

In calibration, each of the threshold parameters (h_{cold}^{thr}, T_{cold}^{thr}, T_{melt}^{thr}, RH_{melt}^{thr} and Pr^{thr}) was discretized to cover the practical, realistic range, and the calibration was then performed in a multiple loop, where each combination of the parameters was tested using the algorithm and a suitable reward function.

The comparison between the algorithm results and FZRA observations can be presented in a 2×2 contingency table (Table A1 in Appendix A), so standard verification measures (e.g., Jolliffe and Stephenson, 2012) can be used. Four candidates were tested to find the appropriate reward function for optimization, namely the Proportion Correct (PC), the Critical Success Index (CSI), the Heidke Skill Score (HSS), and the Symmetric Extremal Dependence Index (SEDI; see Appendix A for the definition of these and other verification measures used in this study). All candidates are positively oriented ~~:~~ so that the higher the value, the better the agreement between the predictions and observations.

When only one measure is used, the CSI is the best one, because strongly biased solutions automatically get worse CSI values. However, all tested reward functions tend to favor biased solutions, either by overestimating (PC, HSS, SEDI) or underestimating (CSI) the total number of FZRA events over all stations and years. This is not desirable as the main interest of the study is in the occurrence climatology of FZRA. Therefore an additional, bias-dependent term was added, and the final form of the reward function was:

$$\mathrm{J} = \mathrm{CSI} - |\log\ \mathrm{B}|, \tag{1}$$

where B is the bias (A7). The logarithm scales B from $-\infty$ to $+\infty$, the best, unbiased value being zero. Therefore biased solutions result a non-zero bias term that is then subtracted from CSI. Adjusted values used in the algorithm are presented in Table 1.

2.5 Climatological analysis

Climatological analysis was performed separately for SYNOP observations, for ERA-Interim in station locations, and for ERA-Interim in the original grid. The following statistics were calculated:

– Total numbers, mean annual numbers, and mean monthly numbers of 6-hourly FZRA events per station or grid cell in 1979–2014. In this analysis one FZRA event occurs when one station or grid cell encounters freezing rain in one time step.

– Spatially averaged annual mean numbers of FZRA events in 1979–2014. Spatial averaging was performed over stations in subgroups, and over all stations. Definition of an event is the same as above.

– Duration of FZRA events separately for station data and for gridded data. In this analysis one FZRA event occurs when one station or grid cell encounters freezing rain in one or in successive time steps. Durations were calculated from a 1-dimensional vector containing the time series of stations or grid cells in a row. Station data with ~~200~~ 293 stations and 53 000 6-hourly time steps comprises a total of ~~28~~ 15 million data points. The gridded reanalysis data consists of 4000 grid cells and 53 000 time steps with altogether ~~210~~ 200 million data points.

– Spatial extent of 6-hourly FZRA events separately for station data and for gridded data. In this analysis one FZRA event occurs when one or multiple stations or grid cells encounter freezing rain simultaneously. The spatial extent was calculated based on the number of impacted stations or, for gridded data, based on the spatial coverage of impacted grid cells over the domain, using an approximative ~~6400~~ 6000 km^2 grid cell size.

After derivation of durations and spatial extents of events, empirical probability distributions (containing events and non-events) were formed~~and plotted against corresponding sorted values of the variables~~.

3 Results

In the following, we first have a look at the refined threshold values in the FMI_{CLIM} algorithm. The performance of the algorithm in predicting FZRA cases is then assessed using the observed weather station data, and finally, the findings concerning climatological features of FZRA are presented. The cross-validation results are based on calibrations in sub-periods, while other validation results are based on the final calibration.

3.1 Cross-validation of calibration

To study the sensitivity of the threshold values to the selection of the calibration period, a cross-validation framework was applied, where the total of 36 years of data was divided into five non-overlapping sub-periods. The calibration was then performed for each of them separately, using Eq. (1). Each sub-period contained a 29-year calibration part and a 7-year validation part so that the validation years were different in all sub-periods. The means calculated over the results from different calibration

periods were used for the climatological analysis of FZRA: they are close to the values that were achieved using the whole 1979–2014 period for calibration.

The calibration changed most of the threshold parameter values only slightly (Table 1), which confirms that (1) the FMI$_{NWP}$ algorithm performs as designed, and (2) no strong biases in mean values of different variables exist in the ERA-Interim data. For example, if the mean temperatures in some of the layers studied were far from the reality, the optimal value would have drifted away from the physically motivated 0°C limit in the calibration. As an exception, the minimum depth of the near-surface cold layer h_{cold}^{thr} was notably altered by the calibration, as the optimal value appeared to be ~~65~~ 69 hPa, which is over 300% larger compared to the original value, 15 hPa. In the lower troposphere these correspond roughly to the depths of ~~400~~ 600 and 130 metres respectively. The observed range of the minimum cold layer depth varies considerably between stations, as shown by Bernstein (2000), being between 100 and 900 metres; 600 metres is thus likely to be more representative than 130 metres over all European stations. The bias-dependent part of the reward function (Eq. 1) was useful in stabilizing the calibration and excluding the less credible combinations of threshold values. Without it, calibration introduced new biases (not shown) either in the annual number of the FZRA cases in subgroups (Fig. 3), or in the total number of cases per station (Fig. 1).

As Table 2 shows, the calibration enhanced most validation metrics, except bias and F. On average, the hit number a was ~~improved by 20%~~slightly improved, and the false alarm b and miss c numbers both were reduced. While H ~~clearly~~ slightly improved, the variability of b was rather large, and the change of F is not statistically ~~significantly~~significant. CSI was improved by ~~20~~8%, HSS by ~~20~~7% and SEDI by ~~5~~2% in calibration, and these changes were statistically significant. The absolute values of CSI and HSS are rather modest, but it is well known (Jolliffe and Stephenson, 2012) that these measures tend to zero when the base rate is very low, as is the case of FZRA. SEDI is designed for the evaluation of rare events and gives much higher values. Still, only one event in five is correctly detected (H $\approx$ 0.20).

3.2 Performance of the freezing rain detection algorithm

3.2.1 SYNOP weather code classification

The algorithm-based (FMI$_{CLIM}$) classifications of weather situations to FZRA events were compared with SYNOP observations of present weather at the validation stations. In Fig. ~~5~~6a, the distribution of observed SYNOP codes, when an event was classified as FZRA, is presented. The distribution shows that only 22% of classified events coincide with SYNOP codes for FZRA (codes 24, 66, and 67). About ~~30~~32% of classifications coincide with codes when no rain of any kind was observed at the SYNOP stations (codes 2 and 10), although reanalysis data did imply rain of at least ~~0.3~~ 0.4 mm/6h (Table 1). Codes associated with light precipitation of other types (drizzle, rain and snow, codes 56, 61, and 71) coincide with FZRA classification, light rain about 6% and other two about 7% each. However, most of these codes occur much more often than codes associated with FZRA. In order to illustrate how the relative numbers of false alarms deviate between different SYNOP present weather codes, the proportion of classified FZRA events in each code is shown in Fig.~~5b~~ 6b. For SYNOP codes associated with FZRA, this proportion can be interpreted as the hit rate H (A5). For SYNOP codes not associated with FZRA, this proportion can be interpreted as the false alarm rate F (A6). Encouragingly, the largest proportions are for codes that are for light FZRA (codes

66, H=2627%), and moderate to heavy FZRA (code 67, H=3431%). The proportion for FZRA observed during past hour (code 24) is somewhat lower (H=109%) but still non-zero. Other clearly non-zero proportions are for codes of freezing drizzle (codes 56, F=6%, and 57, F=68%) and ice pellets (code 79, F=1316%) that physically resemble FZRA.

3.2.2 Number of freezing rain events at weather stations

The annual numbers of FZRA events are reproduced better as spatial averages across the weather stations than at individual sites (Fig. 3 and Table 3). The spatial averages were computed over all the 200 stations in the calibration set and separately for the stations in the coastal, semi-coastal and continental subgroups. FMI_{CLIM} produces marginally better correspondence to observations than FMI_{NWP}, when considering the correlation coefficients (0.93 vs. 0.90) between the predicted and observed, spatially-averaged yearly numbers of FZRA cases per station in all stations. FMI_{CLIM} has a higher correlation also in all subgroups. FMI_{NWP} reproduces the mean values in subgroups better, while standard deviation, which is generally overestimated by both algorithms, is slightly better modelled by the FMI_{CLIM} algorithm. The RMS error of climatological mean is equal the standard deviation of observations, and, encouragingly, the RMS error of all spatial averages is smaller than the standard deviation of observations, implying results are better using FMI_CLIM than just using the climatology. For individual sites, however, the RMS error is slightly worse than the standard deviation of observations.

The total number of FZRA events at each station according to observations was compared with the number of FZRA events according to FMI_{CLIM} (Fig. 45). The comparisons were performed for all the 200 293 stations in the calibration set and separately for the three subgroups of the stations. In each case the root mean squared error (RMS) and the mean error (ME) are were calculated, and the FMI_{CLIM} was modelled as the function of observations using the local polynomial regression method Loess (Venables and Ripley, 2002). The mean number of events calculated with FMI_{CLIM} is almost the same (ME=0.0) with those observed for "all stations" (Fig. 45a). However, the distributions are rather different, the distribution of FMI_{CLIM} result is somewhat symmetric around the mean, but in case of observations the number of stations with small number of FZRA events is higher, and the tail of stations with high frequency of events is much longer. For "all stations" and small numbers of events, FMI_{CLIM} models the average number of events well with some overestimation, and the Loess curve is very near the diagonal line. However, for larger values, the Loess curve is nearly horizontal, implying that FMI_{CLIM} cannot model properly the stations where the large values occur. In the continental group (Fig. 45d), the curve is nearly horizontalhorizontal, or even negatively correlated, for all values. In the coastal group the RMS error is the smallest but there is an overestimation (ME >underestimation (ME < 0), while for the continental group the RMS error is the largest and there is an underestimation (ME <overestimation (ME > 0). Smaller RMS error in coastal areas can be partly explained by the lack of large values that would contribute to RMS.

Spatially the largest biases at the individual validation stations were found in central Europe The spatial averages were computed over all the 293 stations in the calibration set and separately for the stations in the coastal, semi-coastal, and continental subgroups. Because of the uncertainties in the SYNOP weather station observations or in the ERA-Interim reanalysis, calibration had only minor effect on the spatially averaged annual numbers of FZRA events inside the subgroups or over the whole station network (Fig. 1). A coherent, independent of algorithm, and persistent (similarity between different decades, not

shown) 3 and Table 3). Compared to observations, the calibrated and uncalibrated algorithms both underestimate slightly the coastal mean number and overestimate the continental mean number of events, the semi-coastal group being better modelled. Both versions of the algorithm overestimate the standard deviation in all groups and at individual stations. The annual numbers of FZRA events are reproduced better as spatial averages across the weather stations than at individual sites. The RMS error of spatial averages is smaller than the standard deviation of observations, except for the continental group, implying results in general are better using FMI_{CLIM} or FMI_{NWP} than just using the climatology. For individual sites, however, the RMS error is slightly worse than the standard deviation of observations.

An area of underestimation was found in southern Germanyand in Austriacan be seen e.g. in Germany, Czech Republic, and in most stations of Poland, while overestimation mostly happens in the northern and eastern validation stations . As explained above, two reasons why stations were excluded from the validation were the low observed number of FZRA events andthe different definition of observed freezing rain. Most of the stations that were excluded from the validation due to the low observed number of FZRA events are located on coasts. This (Figs. 1 and 4). These biases are spatially homogeneous, independent of algorithm versions (not shown), and do not change in time. The lack of freezing rain, mostly seen in coastal stations, is modelled well, assuming that observations are correct. On the contrary, the difference between observations and algorithm results are most prominent in the eastern stations, which were excluded from validation due to the different definition of observed freezing rain.

3.2.3 Validation of near-surface predictor variables

In addition to the distance to the coastline, orography is likely to affect the occurrence of FZRA (SectSec. 2.1). This suggests that highly variable terrain might cause strong local maxima in the observed occurrence, as the fringe areas of large plains surrounded by mountains are more prone to FZRA. For example, the slopes and valleys surrounding the Great Hungarian Plain might experience cold air damming, a phenomenon which is associated with ice storms in Northern America (Forbes et al., 1987). This phenomenon can also happen in smaller scales, but presumably it cannot be resolved in our work due to the coarse spatial resolution of the reanalysis data. To explore the possible effect of resolution to the accuracy of reanalysis variables, the correlation coefficients between SYNOP observations and ERA-Interim reanalysis surface variables were calculated in the validation stations. Indeed, correlations for near surface temperature and humidity of low altitude stations (0.81 and 0.65 for temperature and humidity respectively) are higher than in high altitudes (0.52 for temperature, 0.44 for humidity). Presumably, the highest stations usually are located in highly variable terrain, and rather large grid cells of the reanalysis do not represent well the observed variability of these stations. No strong differences in observed and ERA-Interim mean values of the variables were found, but the differences are again larger in the highest stations, so that ERA-Interim overestimates, for example, the temperatures by 1.6°C at higher and only 0.3°C at lower altitudes.

3.2.4 Vertical profiles of temperature and humidity

Figure 6 7 shows vertical temperature and humidity profiles of ERA-Interim in the validation stations during the observed and predicted FZRA events. As intended, the FMI_{CLIM} algorithm picks the cases where well-defined melting and freezing layers

exist (Fig. ~~6~~7c). However, as the wider variability ranges in Fig. ~~6~~7a show, FZRA was observed in much more diverse temperature profiles than predicted. The same feature can be seen in the humidity profiles (Fig. ~~6~~7b, d). The FMI$_{CLIM}$ algorithm selects cases where the melting layer humidity is high, even though in reality the precipitation can be formed above the melting layer and thus humidity in the melting layer can be low during the FZRA events. The observed and predicted number of FZRA cases (Fig. ~~6~~7a – e) is ~~7900~~11000, but only ~~1750~~ 2300 events happened simultaneously in observations and in prediction (Fig. ~~6~~7e, f).

3.2.5 Duration and spatial extent of freezing rain events

The probability of the most common FZRA events, those detected during a single 6-hour time step only, is predicted similarly as in observations by both FMI algorithms (Fig. ~~7~~8a). For longer-lasting events the algorithms produce overestimates that increase towards the extreme tail of the durations; at the 10^{-6} probability level the overestimation is about six hours. The spatial extent of FZRA is also overestimated in the extreme tail (Fig. ~~7~~8b) so that ~~FMI$_{NWP}$ overestimates~~ the algorithms overestimate the number of simultaneously impacted stations by ~~20% and FMI$_{CLIM}$ by 50~~about 30% at the 10^{-4} probability level. Additionally the most frequent events, i.e. one impacted station, are slightly underestimated by both algorithms. Durations are modelled better at continental stations than at coastal and semi-coastal stations, and number of impacted stations is modelled almost correctly at coastal stations but poorly at continental stations, with overestimation of almost 100% by FMI$_{CLIM}$ at the lowest 10^{-5} probability level (not shown).

3.3 Climatology of freezing rain in Europe

The spatially averaged ~~annual mean numbers~~ observed annual mean number of FZRA events in different subgroups varies from 0.8 in coastal stations to ~~1.3~~ 1.2 in continental stations as seen in Table 3. Figure 3 shows that the interannual variability of FZRA events is substantial. Even more importantly, the coefficient of variation – standard deviation divided by mean – is large especially in the continental subgroup: there are years with less than 0.5 FZRA events per station, and years with more than ~~three~~ two events on average. Large coefficient of variation may hamper the anticipation and allocation of resources in road maintenance, for example. Weak positive lag-1-year autocorrelations were found in the annual numbers, ranging from 0.20 in semi-coastal to 0.32 at coastal stations, which indicates weak but non-zero predictability of the annual FZRA number based on the number of the previous year.

The spatial distribution of the annual mean number of events (Fig. ~~8~~??a) shows that FZRA is the most frequent in eastern Europe. Large areas in Belarus, Ukraine and Russia encounter ~~3–4~~ 2–3 6-hourly FZRA events per year~~, locally 4–5 6-hourly events~~. The maximum annual number of FZRA cases is situated ~~over The~~ near the Carpathian mountains, where locally over 5 events were found on average. The spatial distribution of maximum durations of the events (Fig. ~~8~~??b) follow qualitatively the mean occurrence distribution in general, but some areas where FZRA is relatively rare, for example the Benelux countries and the Oslo Fjord, have encountered at least one prolonged event. Almost regardless of the latitude, the coastal and marine areas do not experience FZRA as often as the other regions, apparently because warm water bodies effectively prevent the

occurrence of near-surface cold layers~~that are essential for the formation of freezing rain~~. An exception is the northern Baltic Sea, having long ice cover season and relatively frequent FZRA cases.

The FZRA season begins in northern parts of the Fennoscandia already in September, which can be seen in the monthly climatology maps (Fig. ~~??~~10). In that area the phenomenon is experienced most frequently in November. After that, in December–February, the temperatures drop so low that the melting layer seldom forms. Probably for this reason the total number of FZRA events in northern Europe is rather small, even though the season is the longest, lasting until May. In central and especially in eastern Europe the season is shorter but much more intense, so that the annual number of FZRA events is larger than in the northern parts of the continent.

The most widespread (Fig. ~~7~~8b, d) FZRA events at the 10^{-4} probability level covered over 600 000 km^2 and impacted over 10% of the weather stations simultaneously. The most long-lasting events below the 10^{-7} probability level lasted over ~~50~~ 30 hours (Fig. ~~7~~8a, c). It is worth noting that the most long-lasting cases are not necessarily the same as the most widespread events. The proportion of simultaneously impacted stations in the ~~subgroup~~ subgroups varies from 13% (coastal) through 16% (semi-coastal) to 30% (continental stations) at the 10^{-4} probability level (not shown).

4 Discussion

In this paper a freezing rain detection algorithm has been introduced together with a method to calibrate it. After validation the algorithm was applied to a reanalysis in order to construct the European occurrence climatology of freezing rain. This far, no complete gridded climatologies of freezing rain have been presented for Europe in literature. A physically justified, statistically adjusted algorithm which is mainly based on the vertical temperature profile of the atmosphere was used in the study to ensure the credibility of the result. Sub-daily, quality controlled European wide SYNOP weather station data was used in the statistical adjustments. ~~The different definitions of freezing rain in different countries makes derivation of observation-based climatology of freezing rain difficult. The gridded climatology is thus more homogeneous compared to climatologies based only on weather station data.~~

In validation, the gridded meteorological dataset is compared with the point-like surface observations. Each grid cell represents spatial means in the 0.7° resolution, while weather stations represent more local variability of the atmosphere. It is possible that in some cases FZRA has not been observed at a station even though it has occurred nearby. Although hypothetical, this suggests that our estimates, derived from ERA-Interim~~represents~~ , might at least occasionally represent the occurrence of FZRA inside the 0.7° grid cells better than the stations ~~. Each grid cell represents the mean state of variables inside the cell, which by definition can not be the same as in the weather stations representing very local variability of the atmosphere~~do.

4.1 Possible sources of uncertainty

~~Possible~~ Potential sources of uncertainty in the gridded climatology of freezing rain in Europe, besides the detecting algorithm itself, include human errors in observing FZRA, deficiencies in the ERA-Interim reanalysis data, and effects of subgrid-scale orography. These issues are discussed in more detail in the following.

Observing FZRA correctly remains a challenge for observers, as the phenomenon can be easily confused with ordinary, non-freezing rain. Particularly minor cases are difficult to detect for two reasons: firstly, they do not necessarily cause significant ice accretion on structures, which would help the identification of FZRA, and secondly, confusion of FZRA with freezing drizzle or ice pellets might happen, especially by inexperienced observers. Additionally, ~~minor cases~~ short-term events, which are more common than longer ones (Ressler et al., 2012; Cortinas, 2000), might not be recorded in the 6-hourly observations. Short-term events are difficult to predict using spatially and temporally smoothed 6-hourly ~~time interval due to their short duration (Ressler et al., 2012; Cortinas, 2000), which adds uncertainty to the observations of this study~~reanalysis data.

~~To maximize the reliability of the observations, a~~ A large number of SYNOP stations was used in the study, which is believed to average out random errors ~~. Additionally~~ in calculation of spatially or temporally aggregated results, such as mean annual numbers of events in subgroups (Fig. 3). Additionally, the stations having the most complete time series and regular, high-frequency manual observations were included. ~~The strongest observational biases were identified at the eastern stations (~~Still, the effect of human errors can not be totally removed by applying selection techniques to the existing observations. The strongest difference between observations and algorithm results were identified in eastern Europe (algorithm overestimation and/or observational underestimation, as discussed in ~~Sect~~Sec. 2.1 and as seen in Fig. ~~1) , and for that reason they were excluded from the study. Still, the effect of human errors can not be totally removed by applying selection techniques to the existing observations.~~ 4) and in the continental central Europe (algorithm underestimation and/or observational overestimation). Whether the reason to these differences is in the reanalysis data or in the SYNOP observations, remains unclear.

~~Besides the observational uncertainty, a potential source of uncertainty arises from the ERA-Interim reanalysis data .~~ The low-level wintertime temperature inversions in the ERA-Interim reanalysis data are known to be lacking at Arctic latitudes as shown by Serreze et al. (2012). Arguably their result might be valid at least to some extent outside the Arctic. This uncertainty was considered by calibrating the original FMI$_{NWP}$ algorithm instead of using it as such, but apparently the impact of the bias can not be fully compensated by simple adjustments of the threshold values in FMI$_{CLIM}$. In our analysis the mean temperature profile (red line in Fig. 67a) of the ERA-Interim reanalysis do not show a clear near-surface freezing layer below the melting layer, which either indicates problems in the above-surface temperatures of the reanalysis, or highlights the importance of the warm rain formation mechanism of FZRA compared to the melting layer – cold layer mechanism; FMI$_{CLIM}$ is not able to see the former cases.

Locally, near the mountainous regions, a potentially major source of uncertainty is caused by the orography, which is strongly smoothed in the 0.7° resolution of the ERA-Interim. The results may be especially biased in sub-grid scale valleys, where prolonged FZRA events might be caused by trapped cold air mass. It is possible that the optimal value of h_{cold}^{thr}, found here for FMI$_{CLIM}$, differs from the corresponding uncalibrated value because high-elevation or mountainous stations were included in the calibration, and the original, uncalibrated version is mostly used to predict FZRA over the mostly flat-terrain Finland, where smaller h_{cold}^{thr} might work well or well enough.

In the algorithm, two thresholds were used to detect situations favouring subcooling of raindrops~~:~~ , namely, minimum required depth and the maximum allowed temperature of the near-surface cold layer. However, fulfilling both criteria mentioned above does not totally guarantee the liquid phase: a too cold or a too deep cold layer refreezes the hydrometeors, which is not

taken into account in the current version of the algorithm. This could explain, at least partly, the occasional misclassifications to ice pellets.

4.2 Future work

Further exploration of existing data, i.e. observations and the reanalysis, is needed to deepen the knowledge of the phenomenon, including synoptic analysis of the most extreme cases, calculating the precipitation amounts, and studies of other freezing phenomena, i.e., freezing drizzle and ice pellets. ~~Furthermore, impact assessments taking into account the related factors such as the effect of wind (accelerating the accumulation of ice and straining structures by itself), would be informative.~~ In addition to these, the following are the key issues to improve the credibility of the current occurrence results. The list is ordered so that the most important tasks are presented first:

– If the bias structure between the eastern and central European SYNOP stations (Fig. 4) is caused by observational uncertainties, and not by reanalysis or algorithm-dependent uncertainties, identification and rejection of low-quality stations could enhance the calibration and validation processes. Slightly better validation scores between observations and prediction was achieved when eastern stations were excluded from the calibration (not shown).

– Increasing the vertical resolution of the FMI_{NWP} algorithm would be helpful, as small differences in vertical layers easily affect the result (Stewart et al., 2015).

– Validation of the vertical temperature profile of the reanalysis data against observational soundings would allow the division of the uncertainty into method-dependent and data-dependent components.

– The description of the near-surface cold layer in the ~~FMI_{NWP}~~ algorithm could be enhanced by defining a closed range where the parameter h_{cold}^{thr} (or T_{cold}^{thr}) should hit, instead of considering lower (or upper) limits alone. Additionally, separating the moist and melting layers to be independent from each other should be tested~~. Increasing the vertical resolution would be helpful, as small differences in vertical layers easily affect the result (Stewart et al., 2015)~~, as well as using wet bulb temperatures instead of air temperatures as predictors in the algorithm.

– ~~The optimal parameter values could be derived from different calibration periods to have a set of differently calibrated FMI algorithms. The results could then be used to create an ensemble of FZRA climatologies for the purpose of uncertainty analyses or sensitivity analyses~~In addition to the FMI algorithm, new identification methodology could be developed or adopted and tested, including statistical classification methods and more complex but well-performing physical methods, such as ones which more explicitly simulate the melting and freezing of descending hydrometeors (e.g., Ramer, 1993).

– Small uncertainties in the location, in space and time, of the moving precipitation patterns in the reanalysis increase the uncertainty of algorithm-based FZRA detection because of the typical short duration of the FZRA events. To some extent, this uncertainty could be studied by taking into account the preceding and following time steps in the observational records, as the original SYNOP data is 3-hourly. Additionally, preliminary analyses (not shown) indicate that 1-day or 5-day averaging could also enhance the correlation of the results with similarly averaged SYNOP observations.

~~– A more accurate division between the eastern and other European SYNOP stations, following the borders of the countries, would enhance the calibration and validation processes.~~

~~– In addition to FMI algorithms, new identification methodology could be developed or adopted and tested, including statistical classification methods and more complex but well-performing physical methods, such as ones which explicitly simulate the melting and freezing of descending hydrometeors (e.g., Ramer, 1993).~~

~~As shown by Serreze et al. (2012), current reanalyses typically have problems in representing wintertime temperature inversions.~~ ~~Therefore, new~~ New predictor datasets need to be tested when available, for example the ~~ERA5~~ ERA-5 of the ECMWF, which is designed to be the successor of ERA-Interim. The most important ~~criterion~~ criteria when selecting new predictor data would be the accuracy of the vertical temperature structure and high temporal and spatial resolutions~~, since the temperature profile needs to be correct for reliable detection of FZRA~~. Additional observational in situ datasets, such as METAR aviation weather reports and atmospheric soundings, could be used in further development of the FMI algorithm.

As discussed above, neither the observations of FZRA nor methods to predict it are perfect, that is, the ground truth is missing. The methods of then estimating the real base rate of a phenomenon and the verification results of detection are much discussed in social and medical sciences (see, e.g. Lewis and Torgerson, 2012), but are little-known in atmospheric sciences (see Hyvärinen et al. (2015) for the first steps). Ideally, these methods require more than two independent sources of data, for example, different observations and method results. For FZRA this requirement can be difficult to fulfill, as there are not many different sources of observations available and methods are usually developed using all available observations. However, exploring these methods would contribute to the better estimation of the occurrence of FZRA.

5 Conclusions

A method for detecting FZRA in gridded ~~, temporally dense~~ meteorological datasets is presented, followed by a climatological European wide mapping for the occurrence of FZRA. The objective of this paper was to develop an algorithm that is simple enough to ~~avoid memory-intensive computation in the analysis of gridded datasets~~be applicable on spatially, vertically, and temporally coarse public gridded climate datasets such as output from climate models, and on the other hand is physically sensible ~~and sophisticated~~ enough to model the complicated conditions leading to FZRA. The low validation results at station locations indicated that uncertainties related to observations, to the identification method, and to the temporal and spatial resolution of the reanalysis, deteriorate the algorithm-based identification of FZRA events. However, it is not clear which uncertainties are the most important, and it is likely that their relative importance ~~varies~~ can vary in space and even in time.

The freezing rain detection algorithm selected for this study was originally developed in numerical weather prediction. The physically motivated internal thresholds of the algorithm were calibrated using the ERA-Interim reanalysis and SYNOP weather station observations. Values of the thresholds did not change considerably in the calibration process, and the simple calibration did not reveal strong biases in the reanalysis, showing that the original thresholds are already adequate for climatological analysis of freezing rain in ERA-Interim.

According to the algorithm-based analysis of the gridded reanalysis data, freezing rain is more common in central and eastern Europe than in the northern parts and over the coastal regions. The FZRA season begins in September and lasts until May in northern Europe. In central and eastern Europe the season is shorter, beginning in October and lasting until April, but

much more intense, leading to more yearly events (typically ~~2–3~~ 1–2 events/year) than in northern countries (typically 0.5~~–2~~–1.5 events/year). In 1979–2014, the most long-lasting FZRA events lasted over ~~50~~ 30 hours and the most widespread events covered over 600 000 km^2 in Europe. ~~The overall probability of a 50-hour event is 10^{-7} – 10^{-8}.~~

Spatially and temporally coherent information about occurrence of FZRA in Europe has been lacking thus far. The gridded output of this study is a preliminary approach to answer this demand, and as such, the current work can be used as a basis for risk analyses, if the underlying uncertainties are carefully kept in mind. For example, questions such as what year contained the most freezing rain events in continental regions of Europe, can be answered reliably using the current data and method. ~~Station scale~~ Analysing spatially aggregated FZRA results from climate model output with the presented methodology should be feasible as well. Station level analyses, however, require further studies.

6 Data availability

Members of the ECMWF can access the MARS archive for the SYNOP weather station data used in this study. ERA-Interim reanalysis data can be obtained from the public server of the ECMWF. Processed data files and Python code are available on request from the corresponding author.

Appendix A: Verification measures

Results of the comparison of two binary data sources can be presented in a 2×2 contingency table (Table A1). If one of these data sources represents the true values, the cells can be named as follows: a the number of hits, b false alarms, c misses and d correct rejections. In this study, these two data sources are the SYNOP observations and algorithmic classifications of FZRA, and the true values are SYNOP observations. The terminology follows Jolliffe and Stephenson (2012).

The simple measure of performance is Proportion Correct (PC), defined as

$$\mathrm{PC} = \frac{a+d}{a+b+c+d}. \tag{A1}$$

The Critical Success Index (CSI) is similar, but ignores the cell d:

$$\mathrm{CSI} = \frac{a}{a+b+c}. \tag{A2}$$

Many different skill scores have been developed and in this study two of them are used: the Heidke Skill Score (HSS),

$$\mathrm{HSS} = \frac{2(ad-bc)}{(a+c)(c+d)+(a+b)(b+d)}, \tag{A3}$$

and the Symmetric Extremal Dependence Index (SEDI),

$$\mathrm{SEDI} = \frac{\ln \mathrm{F} - \ln \mathrm{H} + \ln(1-\mathrm{H}) - \ln(1-\mathrm{F})}{\ln \mathrm{F} + \ln \mathrm{H} + \ln(1-\mathrm{H}) + \ln(1-\mathrm{F})}, \tag{A4}$$

where the hit rate (H), the ratio of correct FZRA classifications to the number of times the FZRA weather code was observed, is

$$\mathrm{H} = \frac{a}{a+c}, \tag{A5}$$

and the false alarm rate (F), the ratio of false FZRA classifications to the number of times the FZRA weather code was *not* observed, is

$$\mathrm{F} = \frac{b}{b+d}. \tag{A6}$$

Finally, the bias is defined as the ratio of FZRA classifications to the number of times the FZRA weather code was observed

$$\mathrm{B} = \frac{a+b}{a+c}. \tag{A7}$$

Author contributions. Matti Kämäräinen collected and processed the data of this study, designed and implemented the optimization and validation procedures in co-operation with Otto Hyvärinen, and prepared the manuscript with contributions from all coauthors. Otto Hyvärinen designed the cross-validation and validated the classification against SYNOP present weather codes and participated in writing. Simo Neiglick developed the original FMI_{NWP} algorithm used in weather predictions. Jaakko Nuottokari compared ERA-Interim and SYNOP observations in station level. Andrea Vajda, Hilppa Gregow, and Kirsti Jylhä helped with the science, wrote parts of Introduction and commented on the manuscript.

Acknowledgements. This work was partly funded by the European Union's Seventh Programme for research, technological development and demonstration under the RAIN project (Risk Analysis of Infrastructure Networks in response to extreme weather; http://rain-project.eu/; grant agreement N° 608166). The work has also received funding from the ~~Ministry of Employment and the Economy~~ State Nuclear Waste Management Fund in Finland and from the Swedish Radiation Safety Authority through the EXWE project (Extreme weather and nuclear power plants) of the SAFIR2018 program (The Finnish Nuclear Power Plant Safety Research Programme 2015-2018; http://safir2018.vtt.fi).

We thank Sami Niemelä for helping with MARS data retrieval, Tiina Ervasti and Curtis Wood for improving the grammar of the article, and Juulia Lahdenperä for commenting the text. Some of the results were presented in annual meeting of the European Meteorological Society (EMS) in September 2015 by Kämäräinen et al. (2015).

References

Bernstein, B. C.: Regional and Local Influences on Freezing Drizzle, Freezing Rain,and Ice Pellet Events, Weather and Forecasting, 15, 485–508, doi:10.1175/1520-0434(2000)015<0485:RALIOF>2.0.CO;2, 2000.

Bernstein, B. C. and Le Bot, C.: An Inferred Climatology of Icing Conditions Aloft, Including Supercooled Large Drops. Part II: Europe, Asia, and the Globe, Journal of Applied Meteorology and Climatology, 48, 1503–1526, 2009.

Bernstein, B. C., Makkonen, L., and Järvinen, E.: European Icing Frequency Derived From Surface Observations, IWAIS XIII, Andermatt, September 8 to 11, 2009, I (Annex V, III—-17 —- III—-20, 2009.

Bezrukova, N. A., Jeck, R. K., Khalili, M. F., Minina, L. S., Naumov, A. Y., and Stulov, E. A.: Some statistics of freezing precipitation and rime for the territory of the former USSR from ground-based weather observations, Atmospheric Research, 82, 203–221, doi:10.1016/j.atmosres.2005.10.011, 2006.

Bocchieri, J. R.: The Objective Use of Upper Air Soundings to Specify Precipitation Type, Monthly Weather Review, 108, 596–603, doi:10.1175/1520-0493(1980)108<0596:TOUOUA>2.0.CO;2, 1980.

Bourgouin, P.: A Method to Determine Precipitation Types, Weather and Forecasting, 15, 583–592, doi:10.1175/1520-0434(2000)015<0583:AMTDPT>2.0.CO;2, 2000.

Call, D. A.: Changes in Ice Storm Impacts over Time: 1886–2000, Weather, Climate, and Society, 2, 23–35, doi:10.1175/2009WCAS1013.1, 2010.

Carrière, J.-M., Lainard, C., Le Bot, C., and Robart, F.: A climatological study of surface freezing precipitation in Europe, Meteorological Applications, 7, 229–238, doi:10.1017/S1350482700001560, 2000.

Changnon, S. a.: Characteristics of Ice Storms in the United States, Journal of Applied Meteorology, 42, 630–639, doi:10.1175/1520-0450(2003)042<0630:COISIT>2.0.CO;2, 2003.

Cheng, C. S., Auld, H., Li, G., Klaassen, J., and Li, Q.: Possible impacts of climate change on freezing rain in south-central Canada using downscaled future climate scenarios, Natural Hazards and Earth System Science, 7, 71–87, doi:10.5194/nhess-7-71-2007, 2007.

Cheng, C. S., Li, G., and Auld, H.: Possible Impacts of Climate Change on Freezing Rain Using Downscaled Future Climate Scenarios: Updated for Eastern Canada, Atmosphere-Ocean, 49, 8–21, doi:10.1080/07055900.2011.555728, 2011.

Cortinas, J.: A climatology of freezing rain in the Great Lakes region of North America, Monthly Weather Review, 128, 3574–3588, doi:Doi 10.1175/1520-0493(2001)129<3574:Acofri>2.0.Co;2, 2000.

Cortinas, J. V., Bernstein, B. C., Robbins, C. C., and J., S. W.: An Analysis of Freezing Rain, Freezing Drizzle, and Ice Pellets across the United States and Canada: 1976–90, Weather and Forecasting, 19, 377–390, doi:10.1175/1520-0434(2004)019<0377:AAOFRF>2.0.CO;2, 2004.

Cortinas, J. V. J., Brill, K. F., and Baldwin, M. E.: Probabilistic Forecasts of Precipitation Type, in: 16th Conference on Probability and Statistics in the Atmospheric Sciences, https://ams.confex.com/ams/annual2002/techprogram/paper{_}30176.htm, 2002.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, doi:10.1002/qj.828, http://doi.wiley.com/10.1002/qj.828, 2011.

Degelia, S. K., Christian, J. I., Basara, J. B., Mitchell, T. J., Gardner, D. F., Jackson, S. E., Ragland, J. C., and Mahan, H. R.: An overview of ice storms and their impact in the United States, International Journal of Climatology, 2822, 2811–2822, doi:10.1002/joc.4525, 2015.

Editorial: Freezing rain in Gorski Kotar, Journal of Forestry Society of Croatia, 138, 1–2, http://hrcak.srce.hr/file/176384, 2014.

Forbes, G. S., Thomson, D. W., and Anthes, R. A.: Synoptic and Mesoscale Aspects of an Appalachian Ice Storm Associated with Cold-Air Damming, Monthly Weather Review, 115, 564–591, doi:10.1175/1520-0493(1987)115<0564:SAMAOA>2.0.CO;2, 1987.

Groisman, P. Y., Bulygina, O. N., Yin, X., Vose, R. S., Gulev, S. K., Hanssen-Bauer, I., and Forland, E.: Recent changes in the frequency of freezing precipitation in North America and Northern Eurasia, Environmental Research Letters, 11, 1–16, doi:10.1088/1748-9326/11/4/045007, 2016.

Hosek, J., Musilek, P., Lozowski, E., and Pytlak, P.: Forecasting severe ice storms using numerical weather prediction: the March 2010 Newfoundland event, Natural Hazards and Earth System Science, 11, 587–595, doi:10.5194/nhess-11-587-2011, 2011.

Huffman, G. J. and Norman, G. A. J.: The Supercooled Warm Rain Process and the Specifications of Freezing Precipitation, Monthly Weather Review, 116, 2172–2182, 1988.

Hyvärinen, O., Saltikoff, E., and Hohti, H.: Validation of automatic Cb observations for METAR messages without ground truth, Journal of Applied Meteorology and Climatology, p. 150812135039004, doi:10.1175/JAMC-D-14-0222.1, 2015.

Jolliffe, I. T. and Stephenson, D. B.: Forecast Verification : A Practitioner's Guide in Atmospheric Science, Wiley, Hoboken, 2 edn., 2012.

Kämäräinen, M., Jylhä, K., Vajda, A., Gregow, H., Jokinen, P., and Saku, S.: Identifying the freezing rain events in European gridded climate datasets, EMS Annual Meeting Abstracts, 12, http://meetingorganizer.copernicus.org/EMS2015/EMS2015-32.pdf, 2015.

Klima, K. and Morgan, M. G.: Ice storm frequencies in a warmer climate, Climatic Change, doi:10.1007/s10584-015-1460-9, 2015.

Lambert, S. J. and Hansen, B. K.: Simulated Changes in the Freezing Rain Climatology of North America under Global Warming Using a Coupled Climate Model, Atmosphere-Ocean, 49, 289–295, doi:10.1080/07055900.2011.607492, 2011.

Le Bot, C.: SIGMA: System of Icing Geographic identification in Meteorology for Aviation, 11th Conf. on Aviation Range and Aerospace Meteorology, http://ams.confex.com/ams/pdfpapers/81704.pdf, 2004.

Le Bot, C. and Lassegues, P.: Climatology of icing areas derived from ERA40 analysis, 11th Conf. on Aviation Range and Aerospace Meteorology, http://ams.confex.com/ams/pdfpapers/81706.pdf, 2004.

Lewis, F. I. and Torgerson, P. R.: A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic., Emerging themes in epidemiology, 9, 9, doi:10.1186/1742-7622-9-9, 2012.

Marijn, d. H.: Automated discrimination of precipitation type using the FD12P present weather sensor : evaluation and opportunities, Tech. rep., KNMI, R&D Information and Observation Technology, De Bilt, 2007.

Markosek, J.: Severe freezing rain in Slovenia, The European Forecaster, Newsletter, 38–42, doi:10.1111/j.1468-1331.2010.03280.x, http://www.euroforecaster.org/latenews/meteosi.pdf, 2015.

NASA: "Distance to the Nearest Coast", NASA Ocean Biology DAAC, released June 2009, accessed February 29, 2016, http://oceancolor.gsfc.nasa.gov/DOCS/DistFromCoast/, 2009.

Proulx, O. J. and Greene, D. F.: The relationship between ice thickness and northern hardwood tree damage during ice storms, Canadian Journal of Forest Research, 31, 1758–1767, doi:10.1139/x01-104, 2001.

Ramer, J.: An empirical technique for diagnosing precipitation type from model output, Preprints, Fifth Int. Conf. on Aviation Weather Systems, Vienna, VA, Amer. Meteor. Soc., pp. 227–230, 1993.

Rauber, R. M., Olthoff, L. S., Ramamurthy, M. K., and Kunkel, K. E.: The Relative Importance of Warm Rain and Melting Processes in Freezing Precipitation Events, Journal of Applied Meteorology, 39, 1185–1195, doi:10.1175/1520-0450(2000)039<1185:TRIOWR>2.0.CO;2, 2000.

Reeves, H. D., Elmore, K. L., Ryzhkov, A., Schuur, T., and Krause, J.: Sources of Uncertainty in Precipitation-Type Forecasting, Weather and Forecasting, 29, 936–953, doi:10.1175/WAF-D-14-00007.1, 2014.

Ressler, G. M., Milrad, S. M., Atallah, E. H., and Gyakum, J. R.: Synoptic-Scale Analysis of Freezing Rain Events in Montreal, Quebec, Canada, Weather and Forecasting, 27, 362–378, doi:10.1175/WAF-D-11-00071.1, 2012.

Roberts, E., Nawri, N., and Stewart, R. E.: On the storms passing over southern Baffin Island during autumn 2005, Arctic, 61, 309–321, 2008.

Ryzhkov, A., Reeves, H., Krause, J., and Burcham, H.: Discrimination Between Winter Precipitation Types Based on Explicit Microphysical Modeling of Melting and Refreezing in the Polarimetric Hydrometeor Classification Algorithm, ERAD 2014 - THE EIGHTH EUROPEAN CONFERENCE ON RADAR IN METEOROLOGY AND HYDROLOGY, http://www.pa.op.dlr.de/erad2014/programme/ExtendedAbstracts/198{_}Ryzhkov.pdf, 2014.

Serreze, M. C., Barrett, A. P., and Stroeve, J.: Recent changes in tropospheric water vapor over the Arctic as assessed from radiosondes and atmospheric reanalyses, Journal of Geophysical Research: Atmospheres, 117, 1–21, doi:10.1029/2011JD017421, 2012.

Sheppard, B. E. and Joe, P. I.: Automated precipitation detection and typing in winter: A two-year study, Journal of Atmospheric and Oceanic Technology, 17, 1493–1507, doi:10.1175/1520-0426(2000)017<1493:APDATI>2.0.CO;2, 2000.

Stewart, R. E., Thériault, J. M., and Henson, W.: On the Characteristics of and Processes Producing Winter Precipitation Types near $0^{\circ}\mathrm{C}$, Bulletin of the American Meteorological Society, 96, 623–639, doi:10.1175/BAMS-D-14-00032.1, 2015.

Vajda, A.: Impacts of severe winter weather events on critical infrastructure, RAIN WP2 Workshop on Past Severe Weather Hazards, Berlin, 27 February 2015, http://rain-project.eu/wp-content/uploads/2015/03/Vajda_RAIN_Berlin_workshop_FMI_AVajda.pdf, 2015.

Venables, W. N. and Ripley, B. D.: Modern Applied Statistics with S, Springer, New York, 4 edn., 2002.

World Meteorological Organisation: Guide To Meteorological Instruments And Methods Of Observation WMO-No. 8 (2008 Edition, Updated In 2010), http://library.wmo.int/opac/doc{_}num.php?explnum{_}id=3160, 2010.

World Meteorological Organisation: Manual on Codes: Part A - Alphanumeric Codes: International Codes, vol. I.1, WMO Publications, Geneva, Switzerland, 2011.

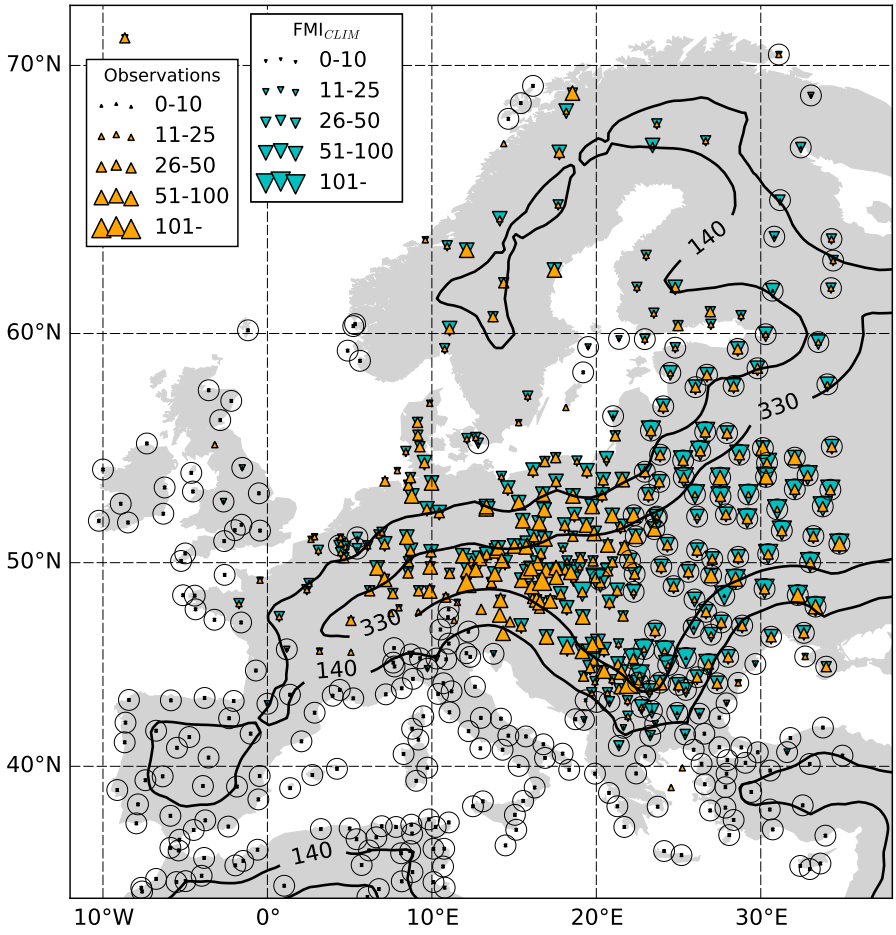


Figure 1. Total number of FZRA cases 1979–2014 in Europe according to 6-hourly SYNOP observations (orange) and the FMI$_{CLIM}$ algorithm (cyan), applied to the 6-hourly ERA-Interim reanalysis. The size of the markings indicates the frequency classes. The distance to the nearest coastline (in km) is shown with black isolines, which divide the stations into coastal, semi-coastal and continental groups. Stations that were excluded from the calibration, validation and further analyses of the FMI$_{CLIM}$ algorithm, are indicated with circles.

```
pLevels = (925, 850, 700)
for all timesteps and stations/gridcells :
    T_cold = T2m
    T_max = max (T_925 , T_850 , T_700)
    if (T_cold ⩽ T_cold^thr and T_max > T_melt^thr and Pr > Pr^thr) :
            p_cold = nearest pressure level (among pLevels) above (p_surf − h_cold^thr)
            moistMeltLayer = any (T [i] > T_melt^thr and RH [i] ≥ RH_melt^thr) ;  i  =  [pLevels ⩽ p_cold]
            if moistMeltLayer exists : FZRA = True
            else : FZRA = False
    else : FZRA = False
```

Figure 2. A pseudocode representation of the FMI algorithm. See text for definitions of symbols and for description of the logic.

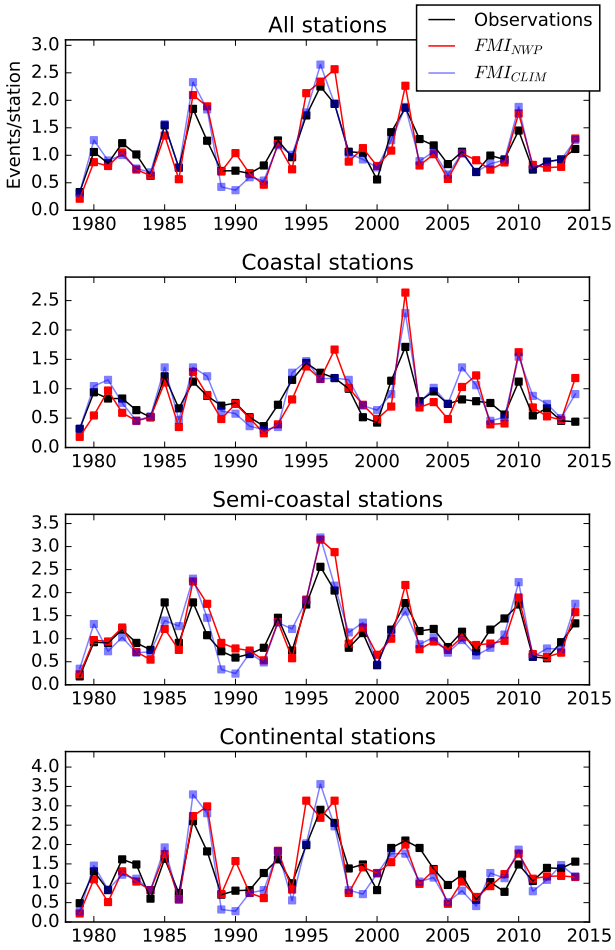


Figure 3. Annual, spatially averaged mean number of FZRA cases per station in all ~~200~~ 293 stations and in groups based on distance to the nearest coastline according to SYNOP observations (black), the FMI$_{NWP}$ (red), and the FMI$_{CLIM}$ (blue) algorithm. Definitions of groups can be seen in Fig. 1. Statistics calculated from the numbers presented here are shown in Table 3. Note the different y-axis scales.

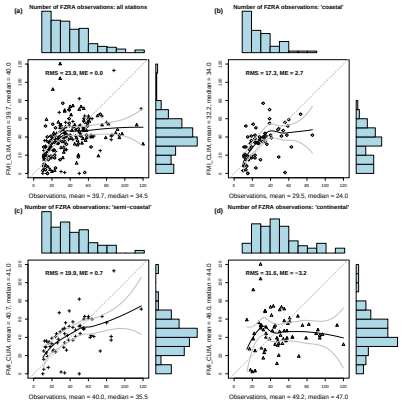


Figure 4. Logarithm of bias B of the FMI_{CLIM} results in different 12-year periods. Blue (red) color shows the overestimation (underestimation) of predicted total number FZRA events compared to observations. Sizes of markings indicate the magnitude of bias.

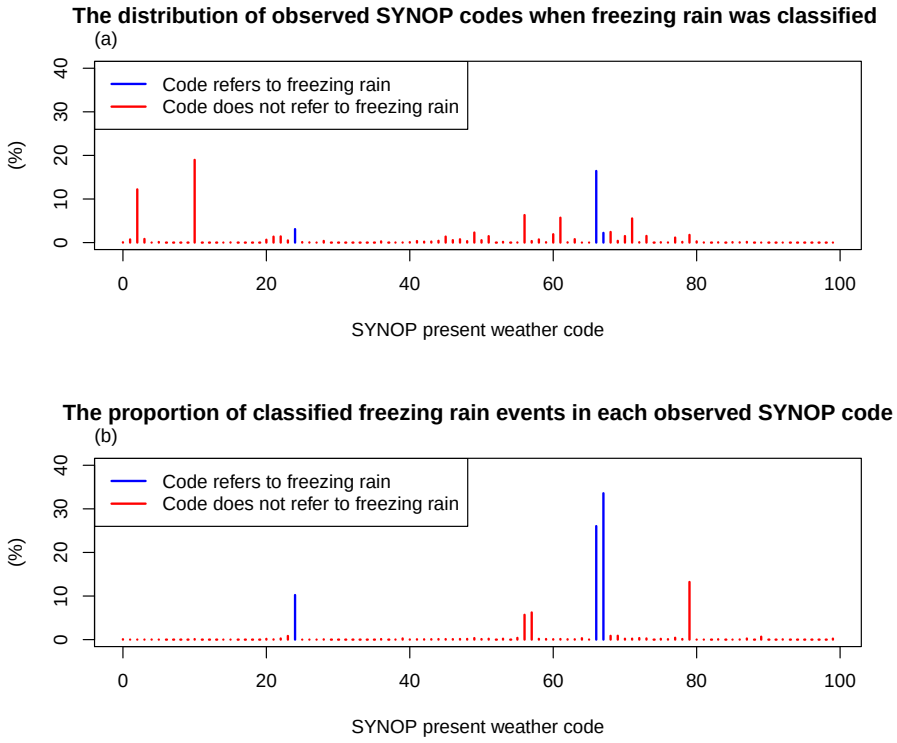


Figure 5. The total number of FZRA events according to observations compared with the number of FZRA events according to FMI_{CLIM} for stations in the calibration set, using (a) all ~~200~~ 293 stations and stations in (b) coastal, (c) semi-coastal and (d) continental groups. The curves superimposed in the scatter plots show FMI_{CLIM} as the function of observations using the Loess method (Venables and Ripley, 2002) together with the 95% confidence interval. Note that plotting symbols for groups in (b), (c), and (d) are used also in (a). The bar diagrams present relative frequency distributions.

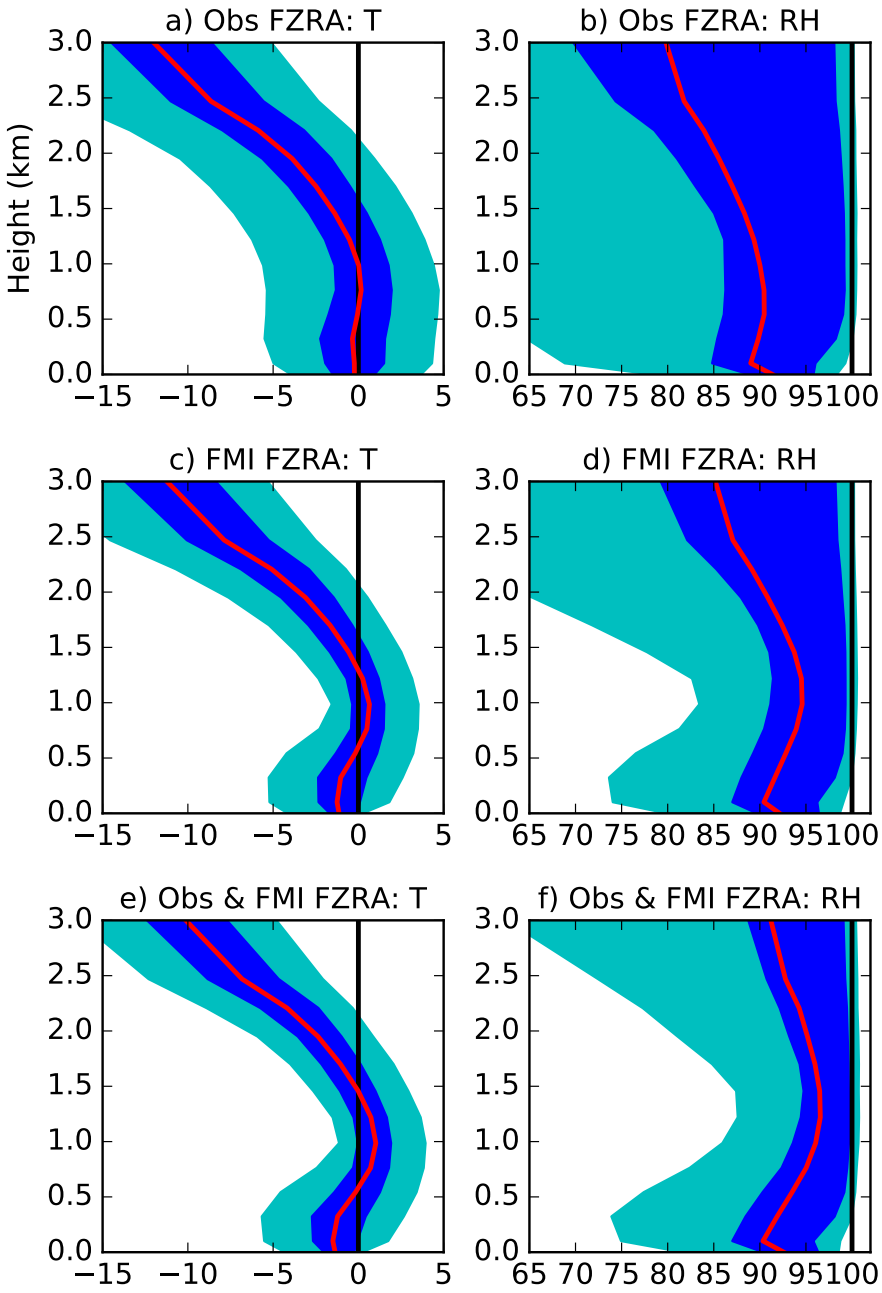


Figure 6. (a) The distribution of observed SYNOP present weather codes when an event was classified as FZRA by the FMI$_{CLIM}$ algorithm. The distribution sums up to 100%. (b) The proportion of cases classified as FZRA in each observed SYNOP present weather code. Each code can have a proportion from 0% to 100%.

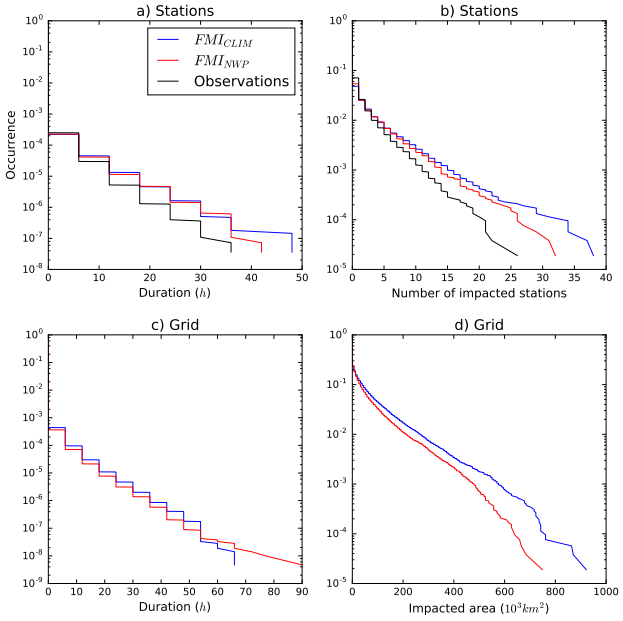


Figure 7. Vertical profiles of temperature (°C, left column) and relative humidity (%, right column) of ERA-Interim at weather station locations. 5%–95% range (cyan), 25%–75% range (blue) and mean (red) are shown. Top row: profiles when FZRA was reported in SYNOP messages (~~7900~~ 11000 events in total). Middle row: FZRA profiles according to the calibrated FMI$_{CLIM}$ algorithm (~~7900~~ 11000 events). Bottom row: profiles where both the FMI$_{CLIM}$ algorithm and observations indicated FZRA (~~1750~~ 2300 events).

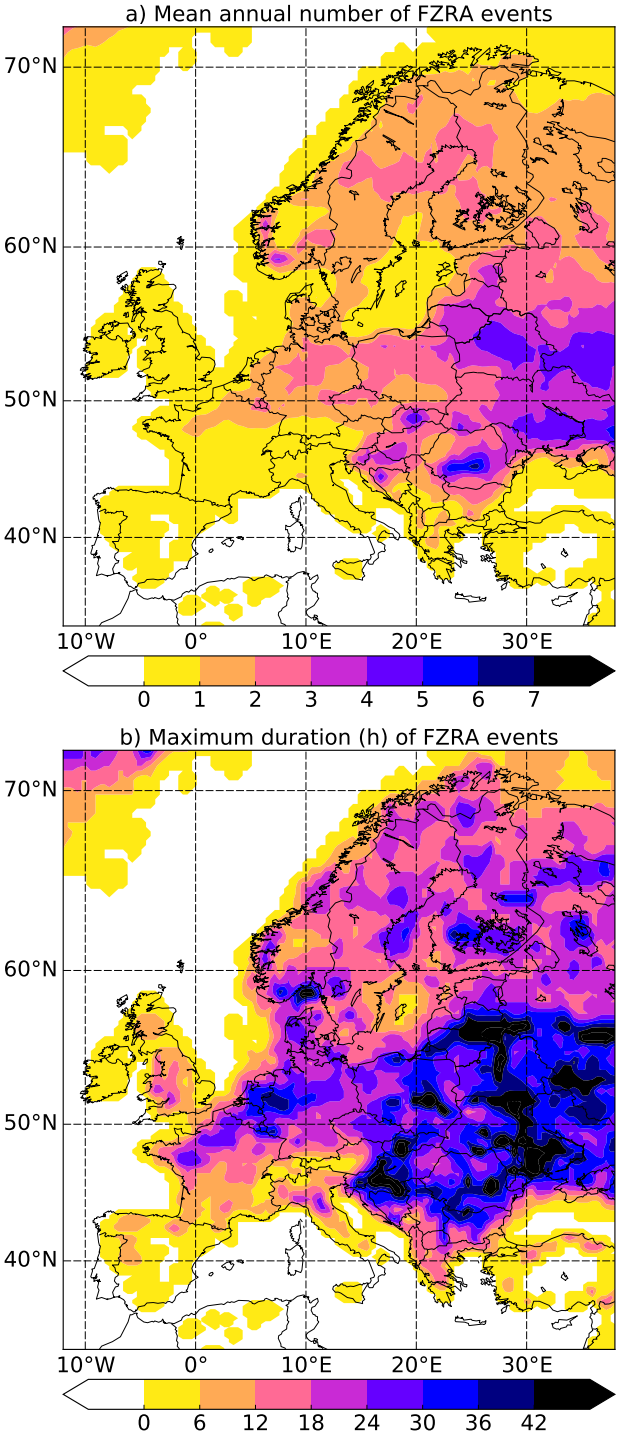
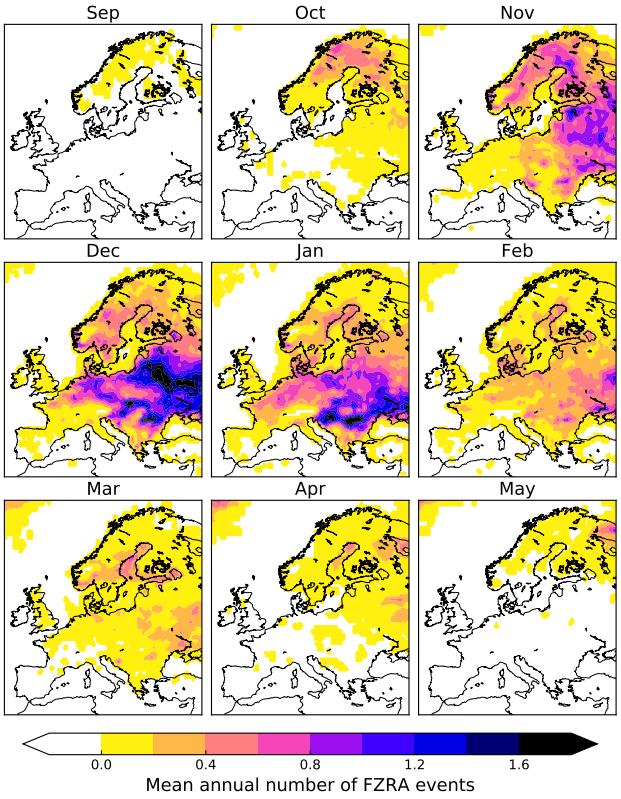


Figure 8. Probability of duration (left column) and spatial extent (right column) of FZRA events at station locations (top row) and in all grid cells (bottom row) according to the detection algorithms (blue, red) and observations (black).



The ~~monthly climatology~~ highest elevations were excluded (gray color) because of ~~FZRA~~ larger uncertainties in ~~1979–2014 according to the FMI$_{CLIM}$ algorithm, applied to the ERA-Interim. The average annual number of 6-hourly~~ FZRA ~~cases is shown~~detection.

Figure 9. (a) Mean annual number of FZRA events and (b) maximum duration of events in the 1979-2014 study period. FMI$_{CLIM}$ algorithm is applied to the ERA-Interim reanalysis data.

The ~~monthly climatology~~ highest elevations were excluded (gray color) because of ~~FZRA~~ larger uncertainties in ~~1979–2014 according to the FMI$_{CLIM}$ algorithm, applied to the ERA-Interim. The average annual number of 6-hourly~~ FZRA ~~cases is shown~~detection.

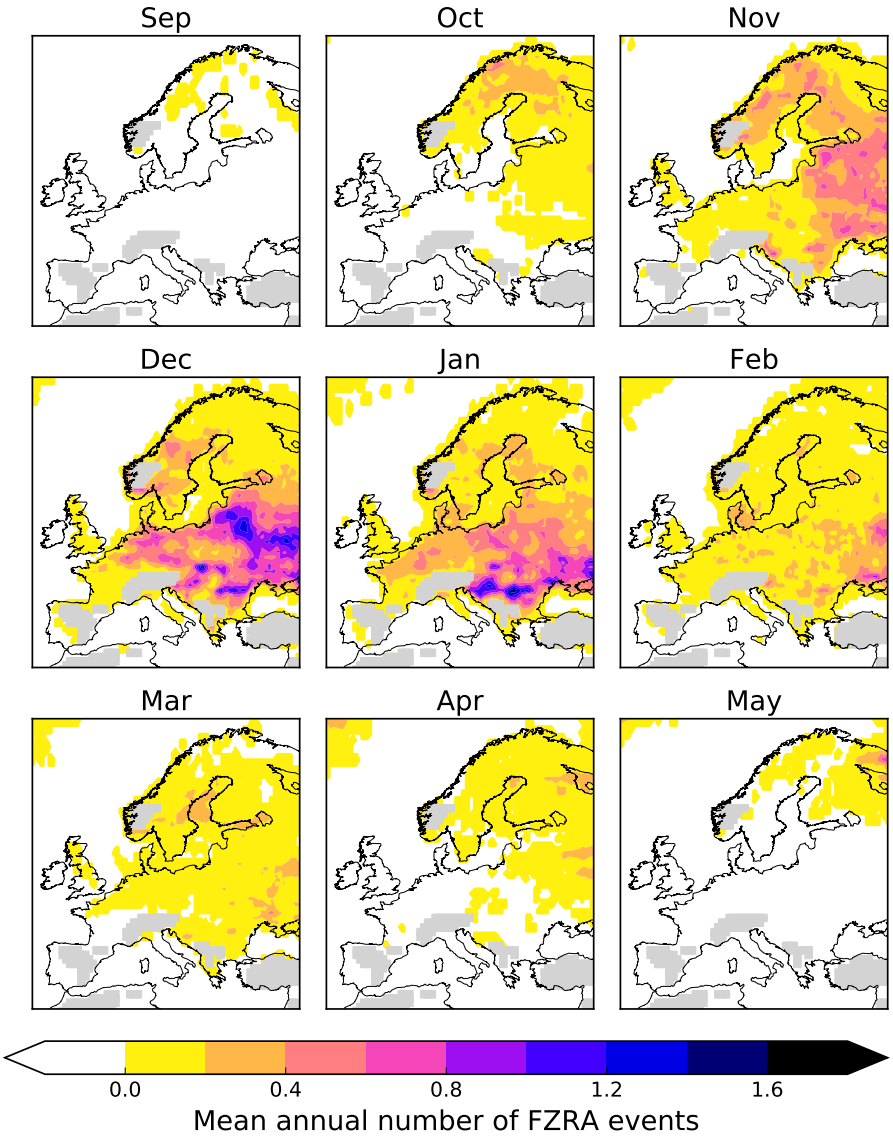


Figure 10. The monthly climatology of FZRA in 1979–2014 according to the FMI$_{CLIM}$ algorithm, applied to the ERA-Interim. The average annual number of 6-hourly FZRA cases is shown. The highest elevations were excluded (gray color) because of larger uncertainties in FZRA detection.

Table 1. Uncalibrated (upper row) and calibrated (bottom), optimal values of threshold parameters in the 29-year calibration periods. Mean values of the optimal values are shown, computed for calibration periods using the sample variance of period values. Mean values are used in the final analysis of the gridded dataset. h_{cold}^{thr} = minimum cold layer depth; RH_{melt}^{thr} and T_{melt}^{thr} = minimum humidity and minimum temperature in the melting layer; T_{cold}^{thr} = maximum cold layer temperature; and Pr^{thr} = minimum surface precipitation rate.

	h_{cold}^{thr} (hPa) >	RH_{melt}^{thr} (%) >	T_{melt}^{thr} (°C) >	T_{cold}^{thr} (°C) <	Pr^{thr} (mm 6h^{-1}) >
FMI$_{NWP}$	15	90	0	0	~~0.04~~ 0.05
FMI$_{CLIM}$	~~65~~ 69	~~86~~ 89	~~-0.58~~ -0.64	~~0.32~~ 0.09	~~0.32~~ 0.39

Table 2. The cross-validation measures and scores in 7-year validation periods when predicted 6-hourly FZRA result is compared against observed 6-hourly events. Mean values and standard errors, computed for validation periods using the sample variance of period values, are shown. See text and Appendix A for definitions of measures and scores.

	FMI$_{NWP}$	FMI$_{CLIM}$	FMI$_{CLIM}$ - FMI$_{NWP}$
CSI	~~0.103 ± 0.005~~ 0.109 ± 0.005	~~0.123 ± 0.007~~ 0.118 ± 0.004	~~0.021 ± 0.004~~ 0.009 ± 0.001
SEDI	~~0.632 ± 0.008~~ 0.65 ± 0.01	~~0.664 ± 0.013~~ 0.66 ± 0.01	~~0.032 ± 0.006~~ 0.014 ± 0.002
HSS	~~0.185 ± 0.008~~ 0.196 ± 0.009	~~0.219 ± 0.012~~ 0.211 ± 0.007	~~0.034 ± 0.006~~ 0.014 ± 0.002
a	~~290 ± 20~~ 430 ± 40	~~340 ± 20~~ 460 ± 40	~~47 ± 11~~ 30 ± 7
b	~~1290 ± 120~~ 1730 ± 150	~~1190 ± 80~~ 1690 ± 160	~~−100 ± 50~~ −40 ± 60
c	~~1280 ± 70~~ 1720 ± 100	~~1230 ± 80~~ 1690 ± 100	~~−47 ± 11~~ −30 ± 7
H	~~0.186 ± 0.009~~ 0.197 ± 0.013	~~0.217 ± 0.015~~ 0.212 ± 0.015	~~0.032 ± 0.008~~ 0.015 ± 0.003
F	~~0.00069 ± 0.00005~~ 0.00063 ± 0.00005	~~0.00064 ± 0.00004~~ 0.00062 ± 0.00006	~~−0.00005 ± 0.00003~~ −0.00011 ± 0.00002
B	1.00 ± 0.05	~~0.97 ± 0.04~~ 1.00 ± 0.07	~~−0.02 ± 0.04~~ 0.00 ± 0.03

Table 3. Statistics calculated from numbers presented in Fig. 3. Correlation coefficient of algorithm results compared to observations (r), mean value ($\overline{x}$), standard deviation (s), and RMS error (RMS) of annual mean numbers of FZRA cases per station averaged over all stations, averaged over groups based on distance to the nearest coastline, and in individual stations are shown.

	FMI_{NWP}				FMI_{CLIM}				Observa
	r	$\overline{x}$	s	RMS	r	$\overline{x}$	s	RMS	$\overline{x}$
All stations	0.90	~~1.10~~ 1.03	~~0.57~~ 0.52	~~0.26~~ 0.25	~~0.93~~ 0.88	~~1.10~~ 1.03	~~0.54~~ 0.51	~~0.22~~ 0.26	~~1.10~~ 1.03
Coastal	~~0.78~~ 0.84	~~0.81~~ 0.71	~~0.48~~ 0.39	~~0.31~~ 0.23	~~0.85~~ 0.86	~~0.89~~ 0.72	~~0.43~~ 0.37	~~0.24~~ 0.20	~~0.82~~ 0.80
Semi-coastal	~~0.89~~ 0.90	~~1.14~~ 1.01	~~0.65~~ 0.53	~~0.31~~ 0.24	~~0.89~~ 0.86	~~1.13~~ 1.01	~~0.63~~ 0.54	~~0.30~~ 0.28	~~1.11~~ 1.02
Continental	~~0.81~~ 0.83	~~1.34~~ 1.26	~~0.76~~ 0.69	~~0.45~~ 0.41	~~0.86~~ 0.83	~~1.28~~ 1.26	~~0.79~~ 0.72	~~0.42~~ 0.44	~~1.37~~ 1.20
Individual stations	~~0.40~~ 0.38	~~1.10~~ 1.03	~~1.61~~ 1.55	~~1.70~~ 1.65	~~0.44~~ 0.38	~~1.10~~ 1.03	~~1.59~~ 1.58	~~1.62~~ 1.66	~~1.10~~ 1.03

Table A1. Contingency table of the comparison between observations and the algorithm. The symbols a-d represent the different number of FZRA events observed to occur in each category.

	Observation	
Algorithm	Freezing rain	No freezing rain
Freezing rain	a (Hit)	b (False Alarm)
No freezing rain	c (Miss)	d (Correct Rejection)