EUMETSAT

NWC SAF

SUPPORT TO NOWCASTING AND VERY SHORT RANGE FORECASTING

Scientific and Validation Report for the Extrapolated Imagery Processor of the NWC/GEO

NWC/CDOP2/GEO/ZAMG/SCI/VR/EXIM, Issue 1.0

22 May 2017

Applicable to

GEO-EXIM-v1.0 (NWC-043)

Prepared by ZAMG

REPORT SIGNATURE TABLE

Function	Name	Signature	Date
Prepared by	R. Goler (ZAMG)		*22 May 2017*
Reviewed by	A. Giannakos (ZAMG) A. Jann (ZAMG)		*22 May 2017*
Authorised by	Pilar Rípodas (AEMET) *SAFNWC Project Manager*		*22 May 2017*

DOCUMENT CHANGE RECORD

Version	Date	Pages	CHANGE(S)
1.0d	15 October 2016	96	Creation (Temporary change bars in the main body highlight changes after STRR 2016)
1.0	*22 May 2017*	95	Addition of ANNEX B: preliminary evaluation results of post-DRR changes in the used displacement field and additional cloud masking were added in Annex B.

Table of contents

List of tables

List of figures

1. INTRODUCTION

The EUMETSAT's "Satellite Application Facilities" (SAFs) are dedicated centres of excellence for processing satellite data, and form an integral part of the distributed EUMETSAT Application Ground Segment (http://www.eumetsat.int). This documentation is provided by the SAF on Support to Nowcasting and Very Short Range Forecasting, NWC SAF. The main objective of NWC SAF is to provide, further develop and maintain software packages to be used for Nowcasting applications of operational meteorological satellite data by National Meteorological Services. More information can be found at the NWC SAF webpage, http://www.nwc-saf.eumetsat.int.

1.1 Scope and purpose of the document

This document is the Validation Report for NWC/GEO Extrapolated Imagery Products (PGE16).

This document contains a description of the validation method and the corresponding results for the above-mentioned product.

1.2 Definitions, acronyms and abbreviations

BT	Brightness Temperature
CDOP	Continuous Development and Operations Phase
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EXIM	Extrapolated Imagery
HRW	High-Resolution Winds
IR	Infrared
LT	Lead Time
MSG	Meteosat Second Generation
NWP	Numerical Weather Prediction
PSS	Peirce Skill Score
ΔPSS	PSS(EXIM) - PSS(Persistence)
PGE	Product Generation Element
SAF	Satellite Application Facility
SAFNWC	SAF to support NoWCasting and Very-Short-Range Forecasting
SEVIRI	Spinning Enhanced Visible and Infrared Imager
VIS	Visible
WV	Water Vapour

1.3 REFERENCES

1.3.1 Applicable Documents

The following documents, of the exact issue shown, form part of this document to the extent specified herein. Applicable documents are those referenced in the Contract or approved by the Approval Authority. They are referenced in this document in the form [AD.X].

For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the current edition of the document referred applies.

Current documentation can be found at the NWC SAF Helpdesk web: http://www.nwc-saf.eumetsat.int.

Ref	Title	Code	Vers	Date
[AD.1]	Proposal for the Second Continuous Development and Operations Phase (CDOP) March 2012 – February 2017	NWC/CDOP2/MGT/AEMET/PRO	1.0d	15/03/11
[AD.2]	NWCSAF Project Plan	NWC/CDOP2/SAF/AEMET/MGT/PP	1.9	15/10/16
[AD.3]	System and Components Requirements Document for the SAFNWC/GEO	NWC/CDOP2/GEO/AEMET/SW/SCRD	1.2	15/10/16
[AD.4]	Interface Control Document for Internal and External Interfaces of the NWC/GEO	NWC/CDOP2/GEO/AEMET/SW/ICD/1	1.2	
[AD.5]	Interface Control Document for the NWCLIB of the SAFNWC/GEO	NWC/CDOP2/GEO/AEMET/SW/ICD/2	1.2	30/10/15
[AD.6]	Data Output Format for the NWC/GEO	NWC/CDOP2/GEO/AEMET/SW/DOF	1.2	
[AD.7]	Architectural Design Document for the NWC/GEO	NWC/CDOP2/GEO/AEMET/SW/ACDD	1.2	
[AD.8]	NWC SAF Product Requirements Document	NWC/CDOP2/SAF/AEMET/MGT/PRD	1.9	15/10/16

***Table 1:** List of Applicable Documents*

1.3.2 Reference Documents

The reference documents contain useful information related to the subject of the project. These reference documents complement the applicable ones, and can be looked up to enhance the information included in this document if it is desired. They are referenced in this document in the form [RD.X].

For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. For undated references, the current edition of the document referred applies.

Current documentation can be found at the NWC SAF Helpdesk web: http://www.nwc-saf.eumetsat.int.

Ref	Title	Code	Vers	Date
[RD.1]	The Nowcasting SAF Glossary	NWC/CDOP2/SAF/AEMET/MGT/GLO		
[RD.2]	User Manual for the Extrapolated Imagery Processor of the NWC/GEO: Science Part	NWC/CDOP2/GEO/ZAMG/SCI/UM/EXIM	1.0	05/05/17
[RD.3]	Algorithm Theoretical Basis Document for "Cloud Products" (CMa-PGE01 v3.2, CT-PGE02 v2.2 & CTTH-PGE03 v2.2)	SAF/NWC/CDOP2/MFL/SCI/ATBD/01	3.2.1	15/07/13

***Table 2:** List of Referenced Documents*

2. GENERAL ASPECTS OF THE VALIDATION APPROACH

2.1 NWCSAF PRODUCTS

EXIM has been applied to the NWCSAF products presented in Table 3.

Product	Abbreviation	Details
SEVIRI thermal infrared	IR3.9, IR8.7, IR9.7, IR10.8, IR12.0, IR13.4	3.9 µm, 8.7 µm, 9.7 µm, 10.8 µm, 12.0 µm, 13.4 µm
SEVIRI thermal water vapour	WV6.2, WV7.3	6.2 µm, 7.3 µm
SEVIRI visible	VIS0.6, VIS0.8, NIR1.6	0.6 µm, 0.8 µm, 1.6 µm
Convective Rainfall Rate v4.0	CRR	
Convective Rainfall Rate from Cloud Physical Properties v1.0	CRPh	
Cloud Type v2.2	CT	
Cloud Top Temperature and Height v2.2	CTTH	
Precipitating Clouds v1.5	PC	
Precipitating Clouds from Cloud Physical Properties v1.0	PCPh	
SEVIRI Clear Air Products Physical Retrieval v2.0 – Precipitable Water	SPhR_TPW SPhR_BL SPhR_ML SPhR_HL	Total Precipitable Water Precipitable water: $P_{surface}$ to 850 hPa Precipitable water: 850 hPa to 500 hPa Precipitable water: 500 hPa to P_{TOA}
SEVIRI Clear Air Products Physical Retrieval v2.0 – Stability Indices	SPhR_KI SPhR_LI SPhR_SHW	K-Index Lifted Index Showalter Index

Table 3: *NWCSAF products used in this analysis*

2.2 VALIDATION DATASET

The validation dataset covers a year from 17th September 2014 to 17th September 2015 for all products except the SEVIRI visible products, where the dataset is from 8th April 2015 to 17th September 2015. This means that the visible products could only be assessed over a full summer and only a part of spring and autumn. The SEVIRI visible channels were only considered between 0600 and 1800 UTC.

The data were taken from the operational NWCSAF v2013+ suite at the ZAMG ("2013+" means: SAFNWC/MSG release v2013, with the product versions stated in **Table 3**, plus the EXIM prototype software). This configuration uses the High-resolution Winds (HRW, v4.0) default model configuration file, with 3-hourly NWP data (ECMWF, 1° resolution across the European area). The geographical area covered is shown in Figure 1.

For this very first release of EXIM, the most important goal of the validation exercise is to prove the validity of the concepts: Is the chosen approach able to reach the minimum threshold accuracy requirement? Which NWCSAF products are amenable to such a method and which are not and should not be offered in the EXIM portfolio? From past experience, it was concluded that the answers to such questions (the behaviour of products) depend on the season rather than on the geographical region or on the specific product release. Hence, the priority for this validation is on capturing the seasonal cycle. In future validation activities, efforts will be made to additionally cover the product versions of the same forthcoming NWC/GEO release for which the validated EXIM version is intended (provided a slight change in the release logic of the project can be implemented; the covered period will necessarily be relatively short as the latest versions naturally become available only shortly before the concerted NWC/GEO release.)

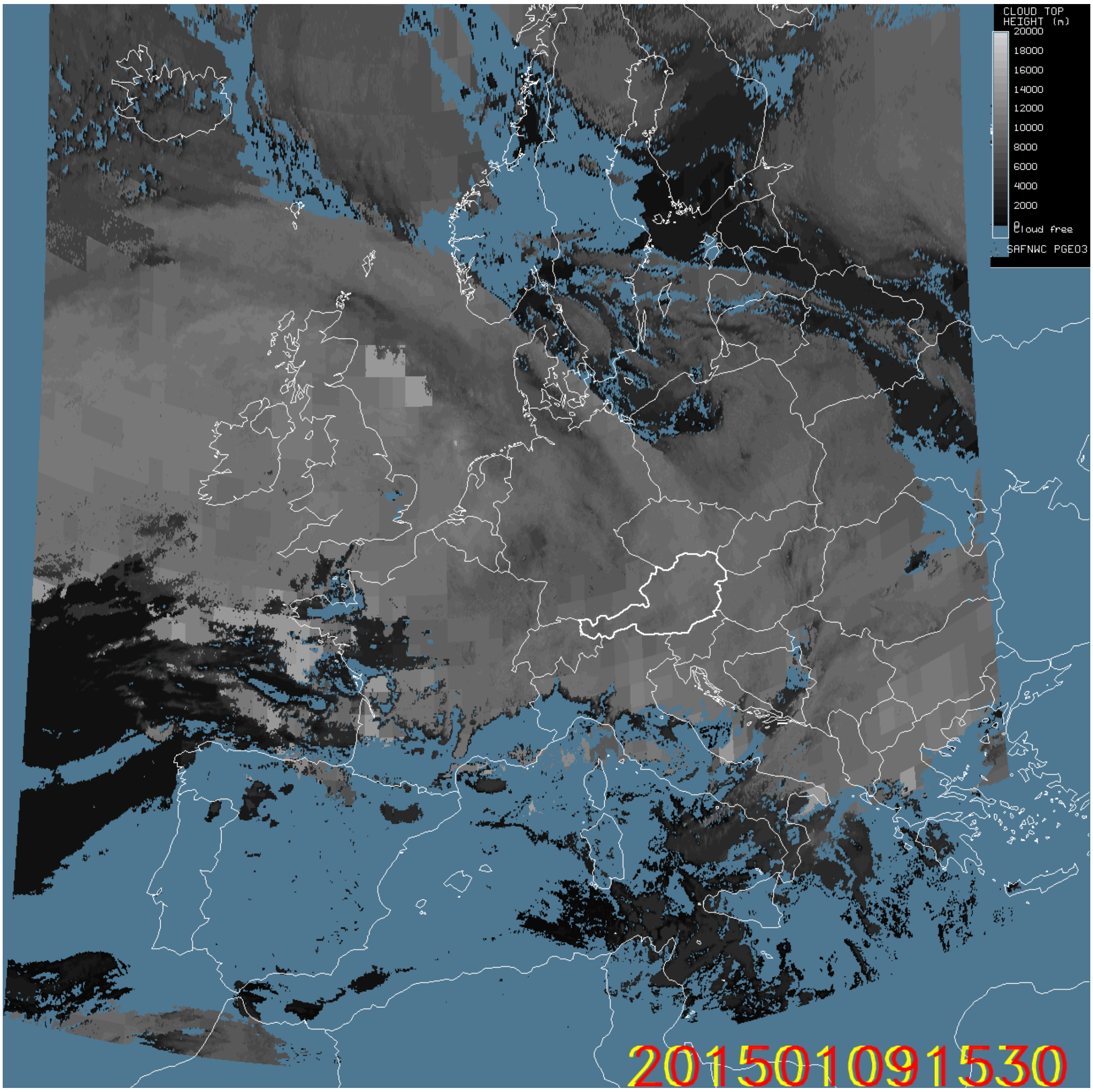

***Figure 1:** Geographical area over which the analysis is performed.*

2.3 Methodological aspects

The SAFNWC Product Requirements Table (PRT) ([AD.8]) defines the threshold accuracy of EXIM as "*on average better than persistence forecast*". The target accuracy is to be "*always better than persistence forecast*". Thus, the validation approach is to compare the EXIM forecast with what was actually observed and verify that the displacement actually made a positive contribution to a skill score.

The EXIM skill scores were logged every two hours starting from 00 UTC (12 times per day); the software produced forecasts at lead times of +15, +30, +45, and +60 minutes. These four times are the times at which the forecasts will be compared.

The Peirce Skill Score, also known as the “true skill statistic”, is a measure of skill obtained by the difference between the hit rate and the false alarm rate of a forecast (see e.g. Wilks[1] 2006). For the 2 × 2 contingency table shown in Figure 2, the hit rate (H) or probability of detection is defined as:

$$H = \frac{a}{a+c} \qquad \text{Range } [0,1].$$

The false alarm rate (F) is defined as:

$$F = \frac{b}{b+d} \qquad \text{Range } [0,1].$$

The Peirce Skill Score (PSS) is thus defined as:

$$PSS = H - F = \frac{ad - bc}{(a+c)(b+d)} \qquad \text{Range } [-1,1].$$

If the PSS is greater than zero, then the number of hits exceeds the false alarms and the forecast has some skill.

The PSS is used here to evaluate the performance of EXIM against the persistence forecast. The analysis of each product is done by choosing a threshold value and assigning a “Yes” (“No”) to each pixel depending on whether its value is above (equal to or below) this threshold. This is done for both EXIM, persistence, and the observed field. The results of each pixel from EXIM and persistence are then compared with those of the observed field and the number of hits (Yes-Yes) and false alarms (Yes-No) counted. It is from these values that the PSS is calculated. The only product where this method cannot be applied is the Cloud Type product which has a (large) number of qualitative categories. For this, a multicategorical variant of the PSS was used, as provided by the *R* package ‘verification’, routine *multi.cont*.

		Observed: Yes	Observed: No	
Forecast	Yes	a	b	a + b
	No	c	d	c + d
		a + c	b + d	n = a + b + c + d

Figure 2: *A 2 × 2 contingency table showing the relationship between counts (letters a, b, c, d) of forecast/event pairs.*

For each lead time (+15, +30, +45, +60 minutes) of every two-hour initialisation, the PSS is calculated over the whole domain twice, using:

[1] Wilks, 2006: Statistical Methods in the Atmospheric Sciences, Elsevier Inc., 649pp.

(i) EXIM as a forecast: PSS(EXIM), and
(ii) Persistence as a forecast: PSS(Persistence).

For a given lead time, this domain average will be referred to as a “case”.

It is possible to interpret this domain average in a number of ways. Consider the case where PSS(EXIM) > PSS(Persistence). Valid interpretations are:

1. EXIM outperforms persistence for all pixels by a small amount
2. EXIM outperforms persistence for a majority of pixels, but is outperformed by persistence for a minority of pixels.
3. EXIM outperforms persistence by a large amount for a minority of pixels, but is outperformed by persistence by a small amount for the majority of pixels.

In order for EXIM to be considered useful, it would be desired for it to produce skilful results for the majority of pixels in the domain (i.e. case 1 and 2). However, the current analysis cannot completely exclude case 3, and is a subject for future work.

In the following analysis, the success of EXIM over persistence will be judged on the number of cases in which EXIM outperforms persistence, and on the amount by which EXIM outperforms persistence. It can be anticipated that EXIM will excel in situations where the flow results in advection of the feature in question (e.g. cloud for the IR channels). However, in cases where cloud features are stationary despite the wind field, such as in the case of lee clouds, or back-building convective elements, persistence will excel. Thus, whether EXIM or persistence produces a higher PSS will depend upon how dominant the above mentioned features are at a given time.

2.4 THERMAL CHANNELS: INFRARED

The SEVIRI thermal infrared channels consist of the spectral bands centred on the wavelengths 3.9 µm, 8.7 µm, 9.7 µm, 10.8 µm, 12.0 µm, and 13.4 µm.

The performance of EXIM compared with persistence for each thermal infrared channel is shown in Figure 3 to Figure 8. Each figure shows the performance for a specific channel over a range of threshold temperatures (individual panels) which correspond to brightness temperatures (BT). Low BT corresponds with high levels in the atmosphere. The data are separated according to season, and a sum for the whole year is also shown. The green columns represent the percentage of cases where PSS(EXIM) > PSS(Persistence) and the red columns represent the percentage of cases where PSS(EXIM) ≤ PSS(Persistence).

In general, EXIM outperforms persistence for the majority of cases. When analysing the results in detail it is evident that the performance of EXIM is better in summer and poorer in winter. In particular, the performance of EXIM is relatively poor for the highest threshold temperature of 280 K for all channels and for the winter of 270 K. The reason for this is that at higher threshold temperatures, surface features (e.g. snow) will be detected, which should remain fixed. In winter, a given threshold temperature will refer to a lower height level than in summer. For example, a BT of 270 K in summer may correspond to a height of 4.5 km on a hot day, but correspond to the surface in winter. As the EXIM algorithm would result in surface features being translated, persistence will naturally outperform EXIM.

The performance of EXIM relative to persistence with lead time shows a decreasing trend in all panels. This means that although EXIM may outperform persistence at lead times of +15 and +30 minutes, persistence will outperform EXIM at later lead times (e.g. Figure 3, 270 K, Winter). This may seem counterintuitive, since one typically associates a wind field with cloud features propagating with the wind. Thus as the lead time increases, the distance of the cloud from its starting point should increase. In such a case, PSS(Persistence) should continually decrease with lead time and PSS(EXIM) should increase, i.e. the trend shown in the results should be reversed with the green columns getting larger for greater lead times. That this is not the case can be explained by several reasons, in order of importance:

1. The decreasing trend is greatest for the highest threshold temperatures. As mentioned earlier, at higher threshold temperatures the detection of surface features becomes more problematic. Persistence will thus outperform EXIM at greater lead times.
2. The decreasing trend can also be seen in the results of the lower threshold values (e.g. Figure 4, 230 K). In these cases the presence of lee clouds, which typically occur in the upper layers of the troposphere, have an effect. As EXIM continues to incorrectly translate the stationary cloud, the error in the cloud’s position between EXIM and persistence increases with increasing lead time.
3. EXIM determines the future movement of pixels based on the trajectories at the initial time. As weather systems generally are non-stationary, their movement will result in changes to the trajectories. Consequently, the trajectories used at the initial time will generally not correspond to the trajectories after an hour, with the difference becoming larger for greater lead times. An example where this may have an effect is in the case of fast moving cold fronts and their associated low pressure systems, typical features over northern and central Europe, particularly in winter.

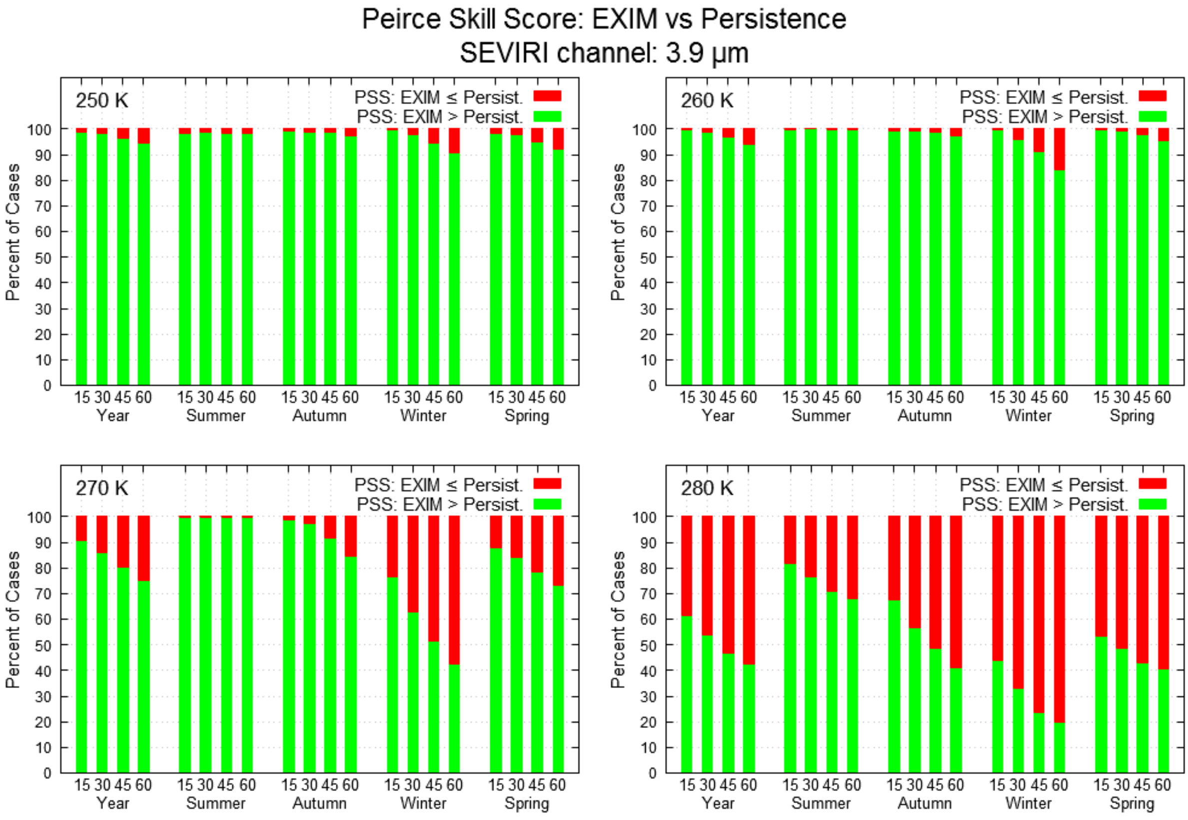

Figure 3: *SEVIRI channel 3.9µm. Histogram showing the percentage of occurrences where the Peirce Skill Score (PSS) of EXIM exceeds (green), or is less than or equal to (red) that for persistence. Each panel corresponds to a threshold value shown in the upper left. Each group of bins represent the data according to season along with a sum over the year, and is separated into the lead times (+15, +30, +45, +60 minutes).*

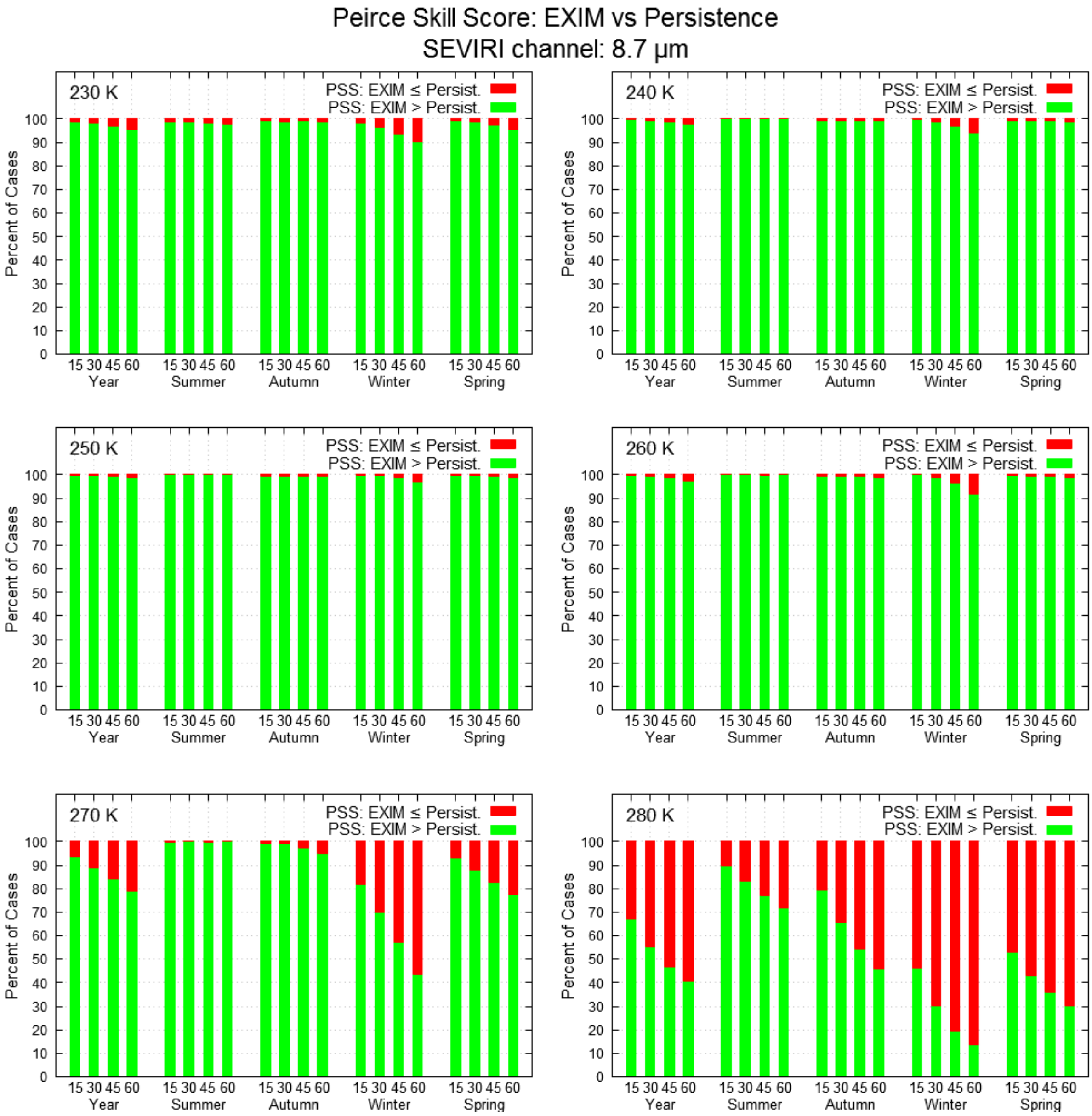

Figure 4: *As in Figure 3 but for SEVIRI channel 8.7µm.*

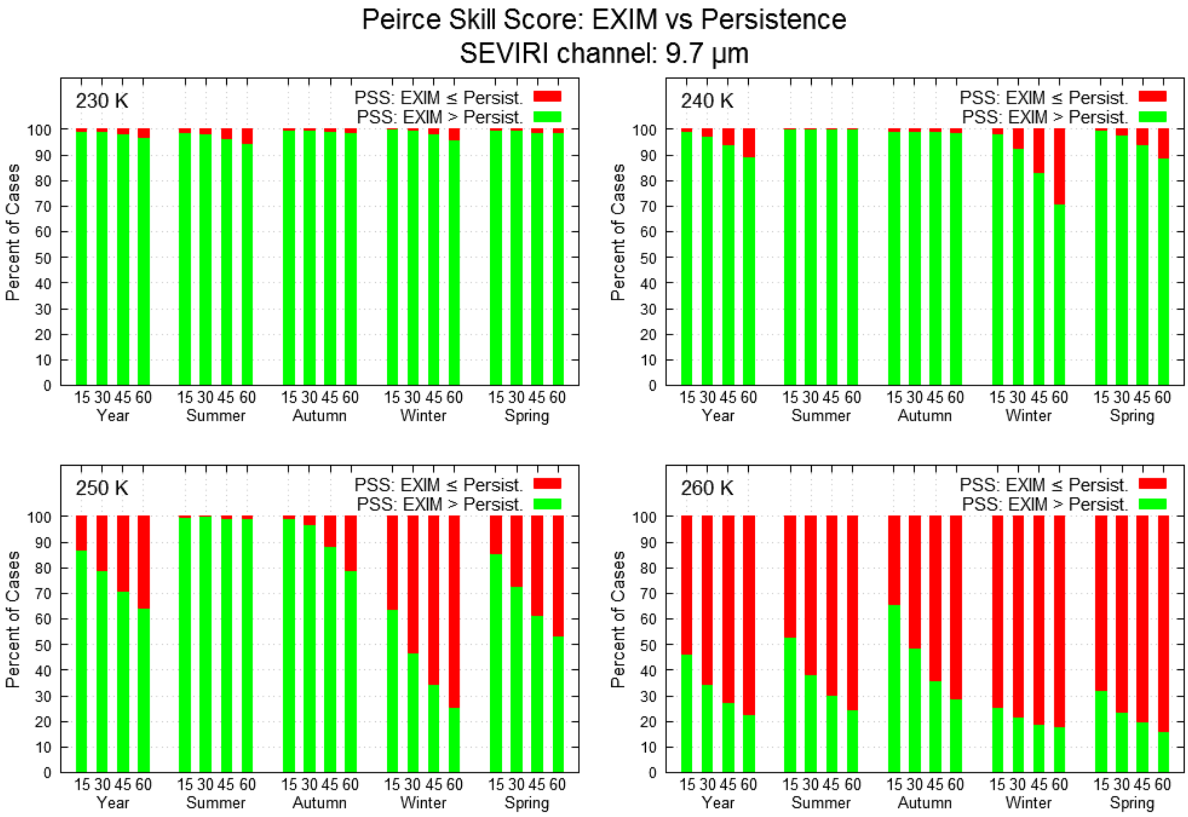

Figure 5: *As in Figure 3 but for SEVIRI channel 9.7µm.*

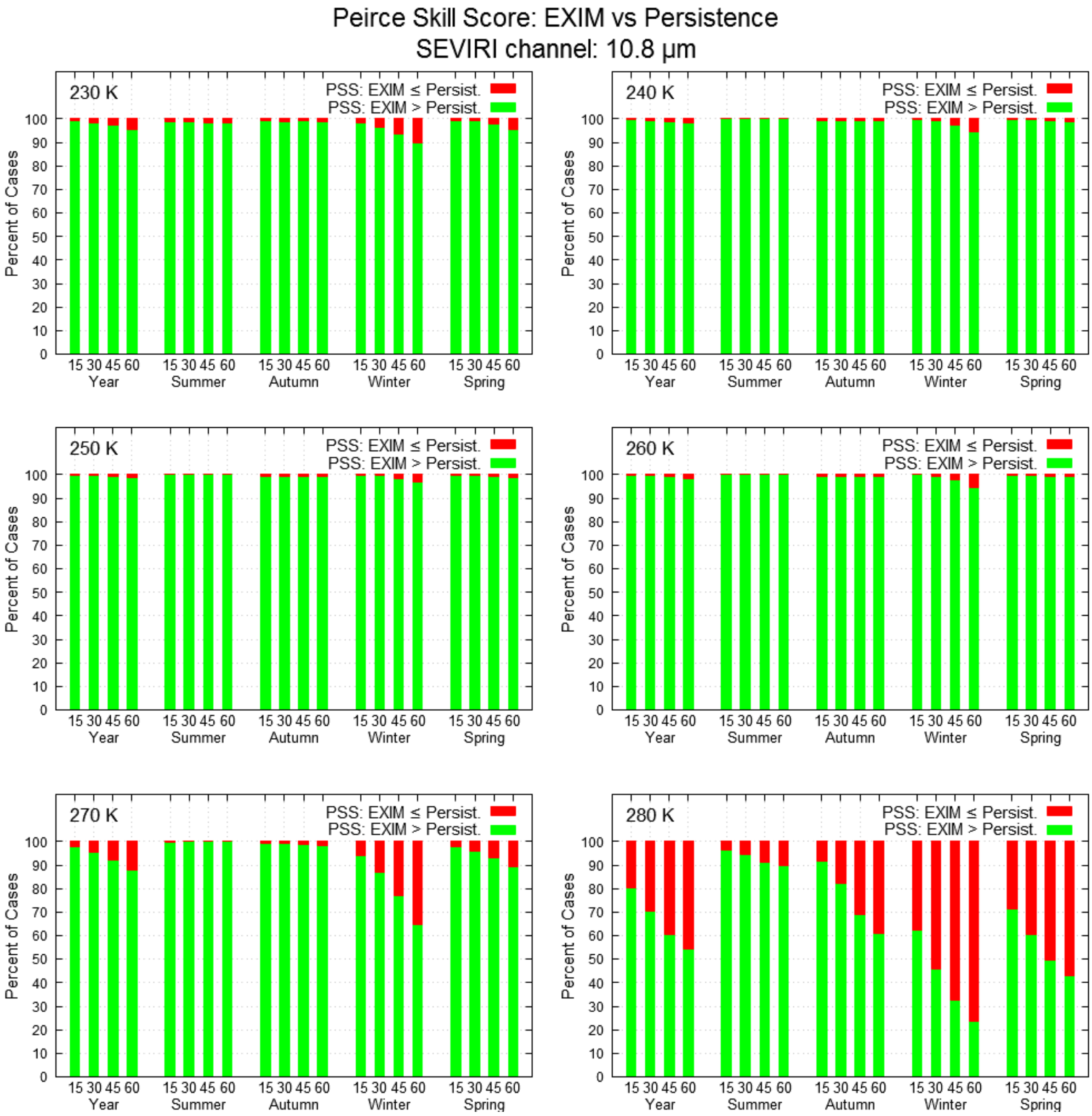

Figure 6: *As in Figure 3 but for SEVIRI channel 10.8µm.*

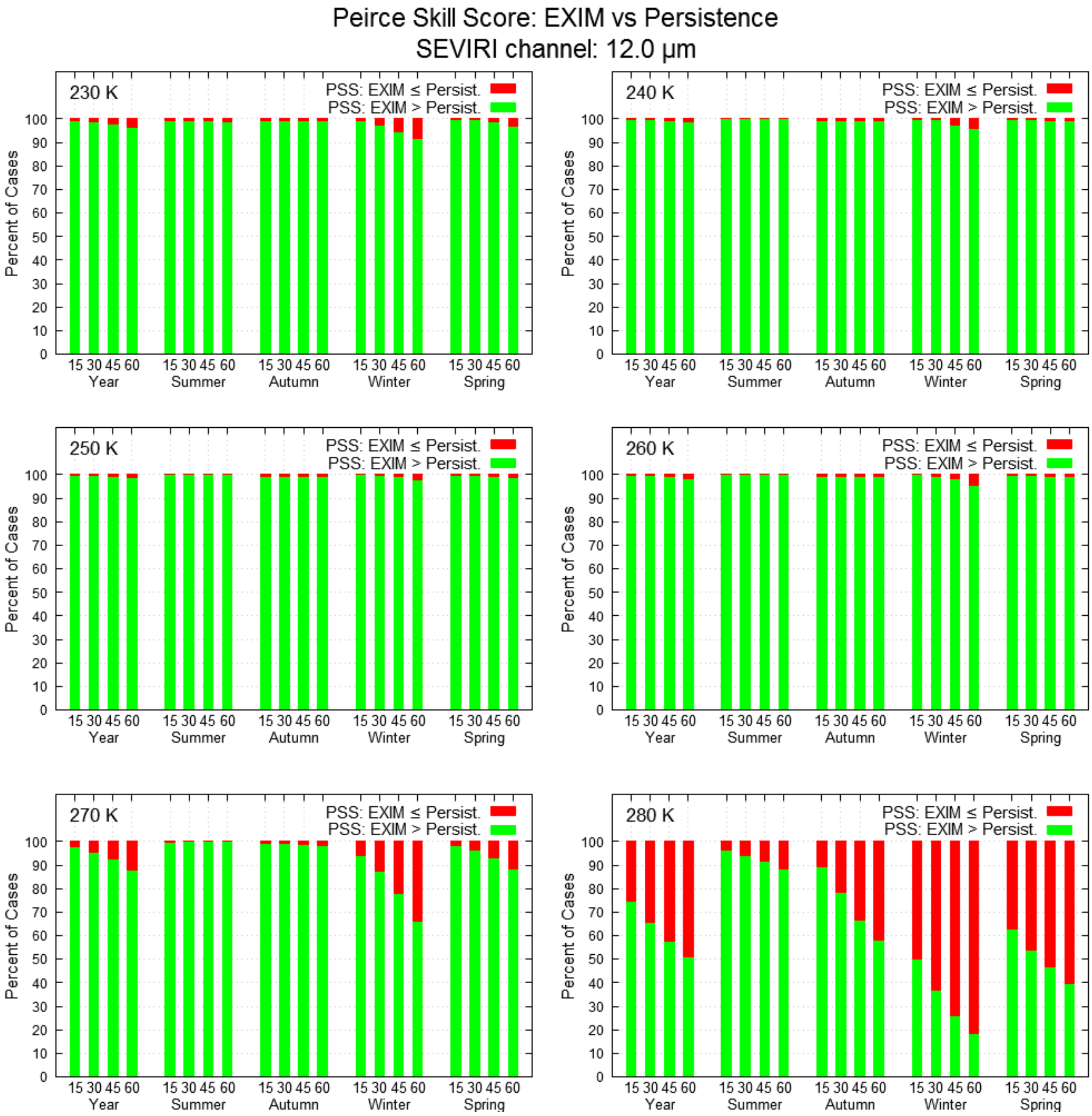

Figure 7: *As in Figure 3 but for SEVIRI channel 12.0µm.*

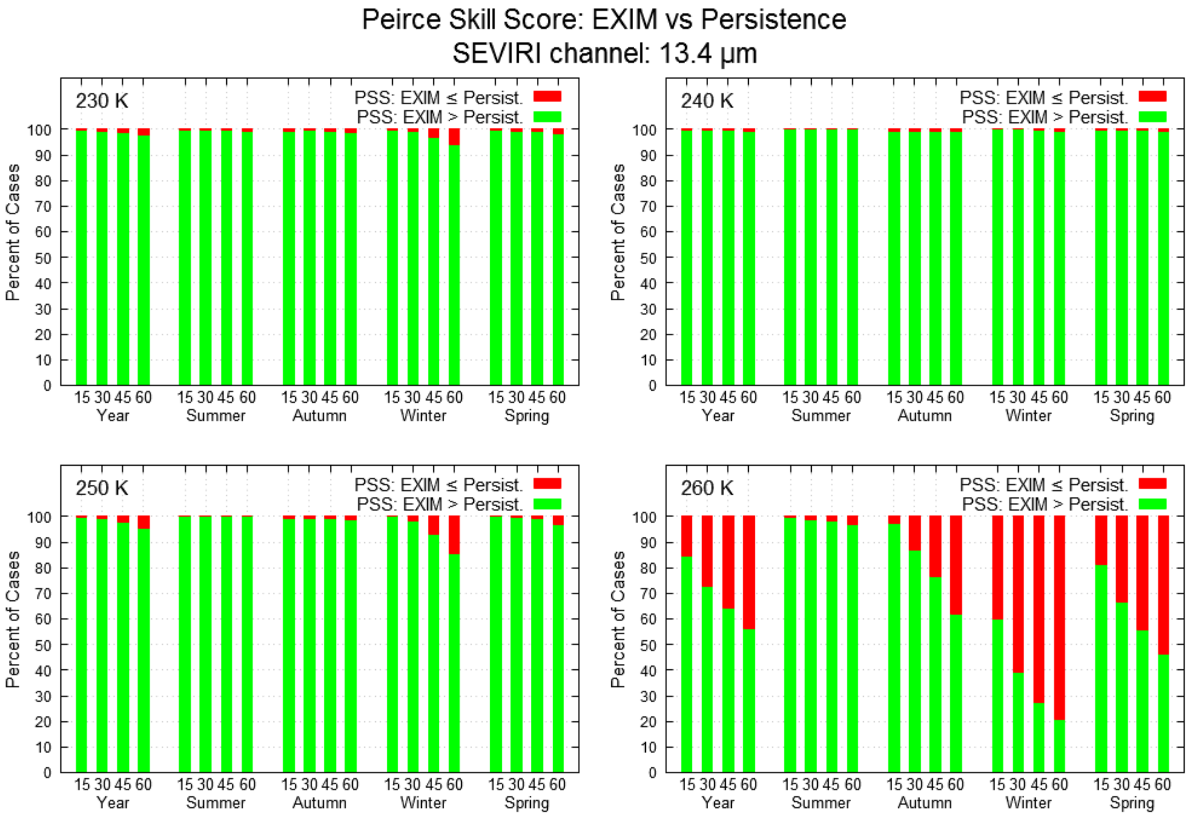

Figure 8: *As in Figure 3 but for SEVIRI channel 13.4µm.*

Figure 9 shows a time series of ΔPSS = PSS(EXIM) - PSS(Persistence) for the SEVIRI channel 10.8 µm and a threshold value of 240 K. The values of ΔPSS at the four lead times (+15, +30, +45, +60 minutes) are represented by a curve (blue, green, yellow, red, respectively) displaced vertically and plotted relative to ΔPSS = 0 for each curve. The corresponding curves for the other channels are similar and are not shown. As lead time increases, ΔPSS becomes smaller. In most cases where ΔPSS is negative, lee cloud was present, whereby fast flowing winds across mountain ranges lead to stationary cloud development downwind thereof. An example of this occurred in the period 23-24 December 2014, with lee cloud development on the southeastern side of the Alps (Figure 10 left panel).

In a few cases, a negative ΔPSS could be attributed to the presence of stationary (back-building) deep convection, whose continual development upstream of the main updraft results in the convective cloud remaining stationary within the wind field. An example of this occurred on 4 December 2014 with deep convection over the Adriatic Sea (Figure 10 right panel).

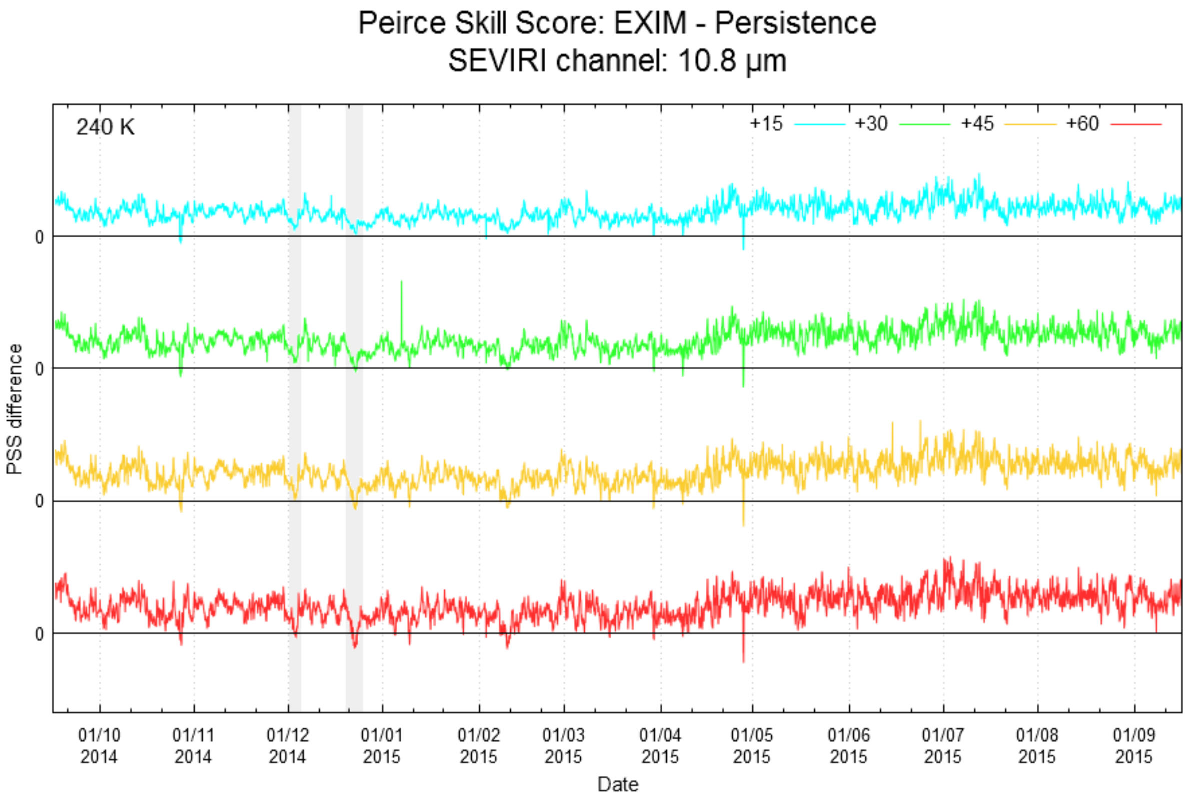

Figure 9: *Time series of ΔPSS = PSS(EXIM) – PSS(Persistence) for SEVIRI channel 10.8 µm and a threshold value of 240 K. The curve for each lead time (+15 min = blue, +30 min = green, +45 min = yellow, +60 min = red) is plotted relative to its y=0 axis. For reference, the distance between two y=0 lines corresponds to a value of ΔPSS = 0.5. The two grey columns represent the dates of the satellite images shown in Figure 10.*

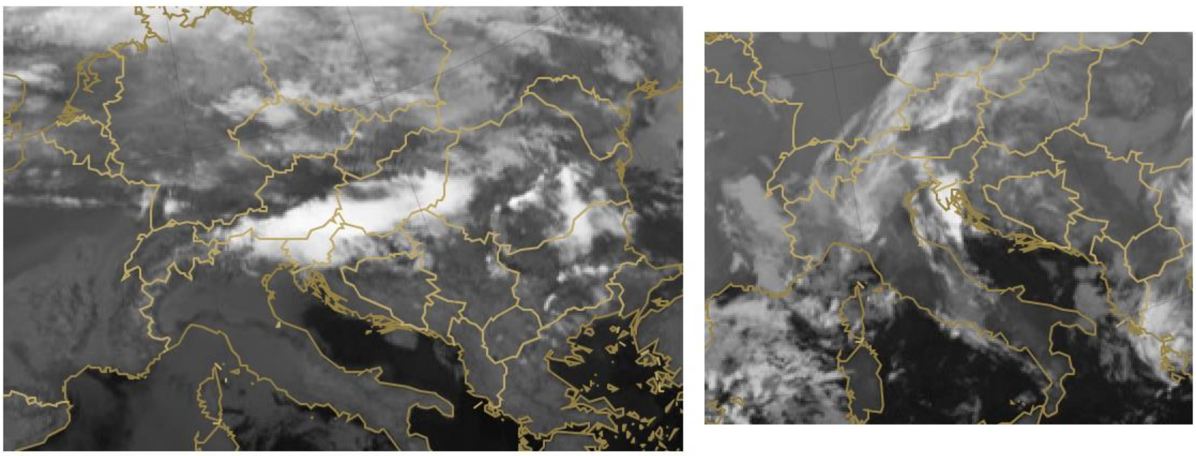

Figure 10: *Left panel: Infrared image on 23 December 2014 0900 UTC showing lee cloud south east of the main Alpine ridge. Right panel: Infrared image on 4 December 0100 UTC showing a back-building deep convective element over the Adriatic Sea. Images courtesy of the ePort platform from EUMeTrain (www.eumetrain.org).*

The previous plots showed the percentage of cases in which EXIM outperforms persistence. However, they do not show by how much EXIM outperforms persistence. In order to show this, the distribution of ΔPSS values is plotted for the two lead times, +15 minutes (dark blue) and +60 minutes (light blue) for each of the channels and thresholds in Figure 11 – Figure 16. In order to have a sufficient number of cases for the distribution, the entire year of data was used.

Except for the highest threshold values, as mentioned previously, EXIM performs better than persistence, as indicated by the peak of the distribution occurring at positive values of ΔPSS. On average, EXIM provides an improvement over persistence in the PSS of approximately 0.1. The majority of the plots display a greater peak and a narrower spread at +15 min compared with +60 min lead time. This shows that at shorter lead times, EXIM is consistently better than persistence, while for later lead times, there do exist casts where persistence is better.

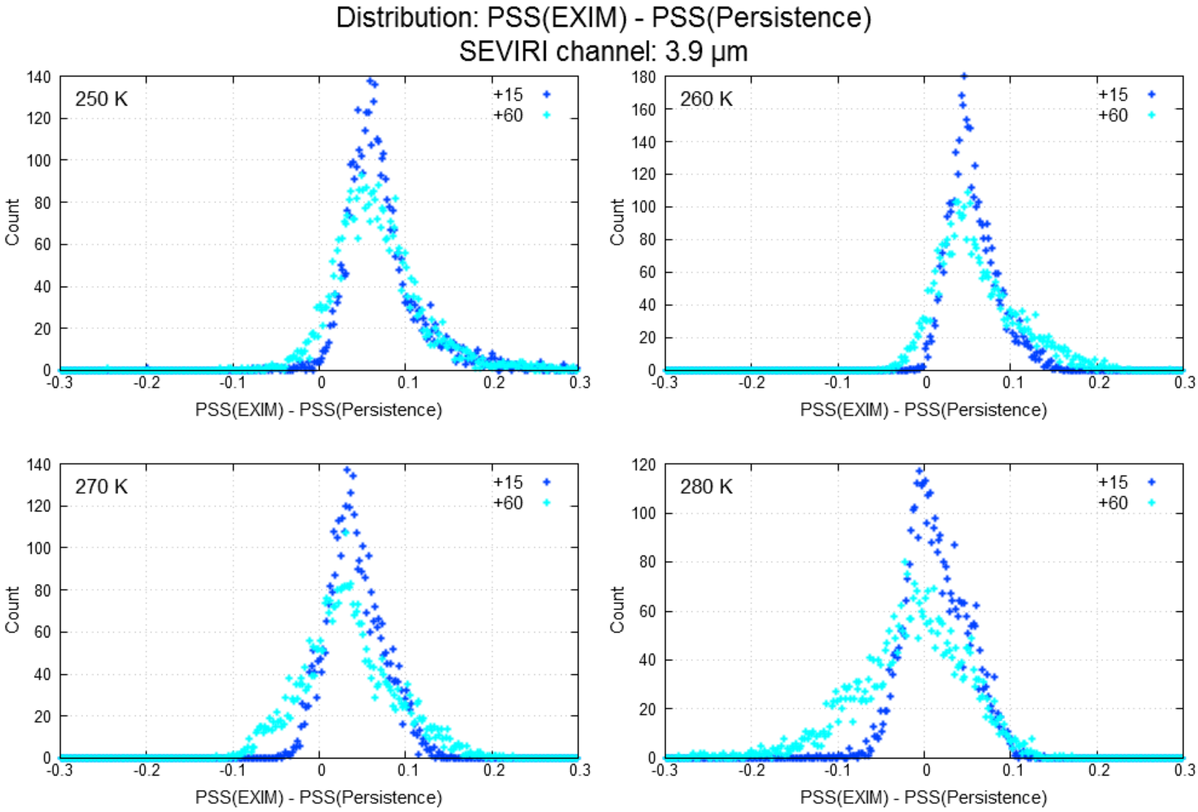

Figure 11: *SEVIRI channel 3.9µm. Histogram showing the distribution of ΔPSS at lead times of +15 minutes (dark blue) and +60 minutes (light blue). The entire year of data is used. Each panel corresponds to a threshold value shown in the top left.*

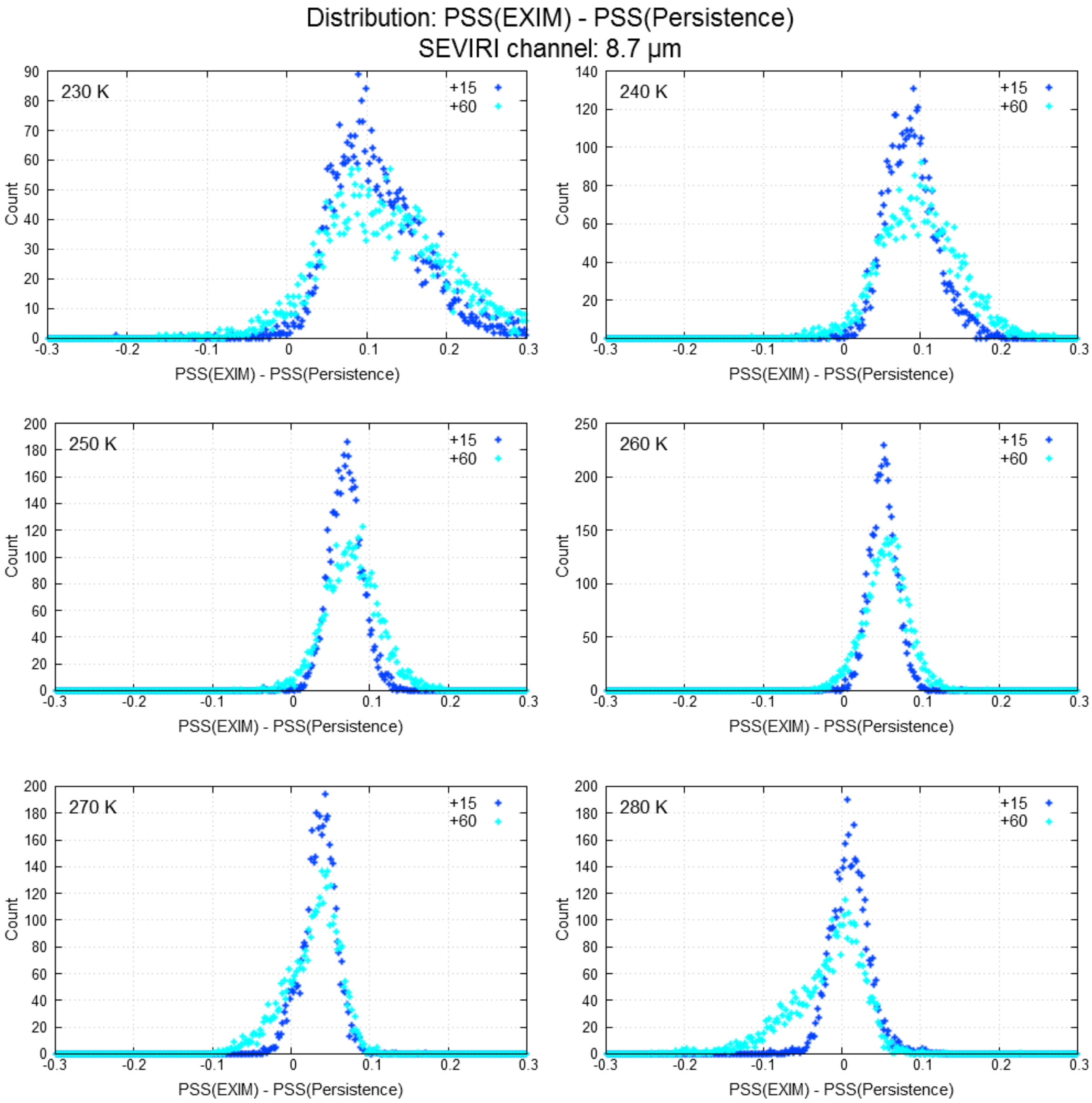

Figure 12: *As in Figure 11 but for SEVIRI channel 8.7µm.*

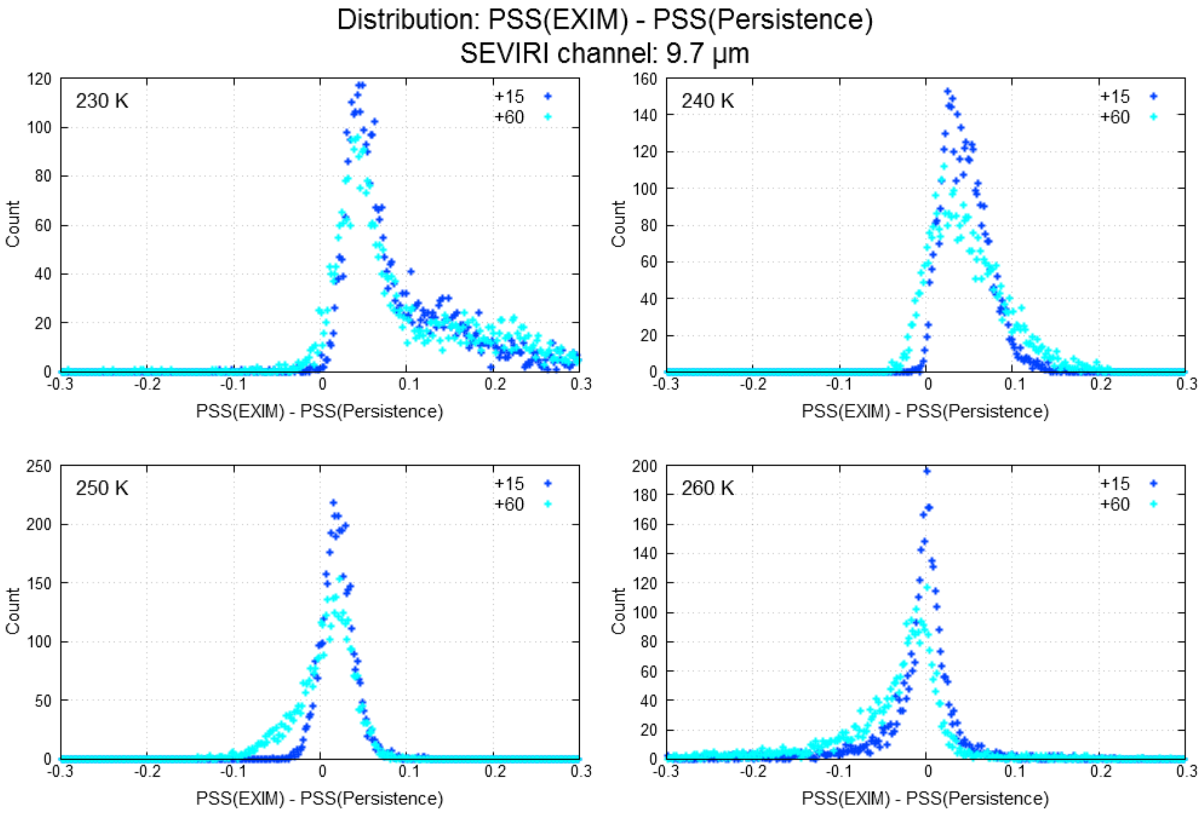

Figure 13: *As in Figure 11 but for SEVIRI channel 9.7µm.*

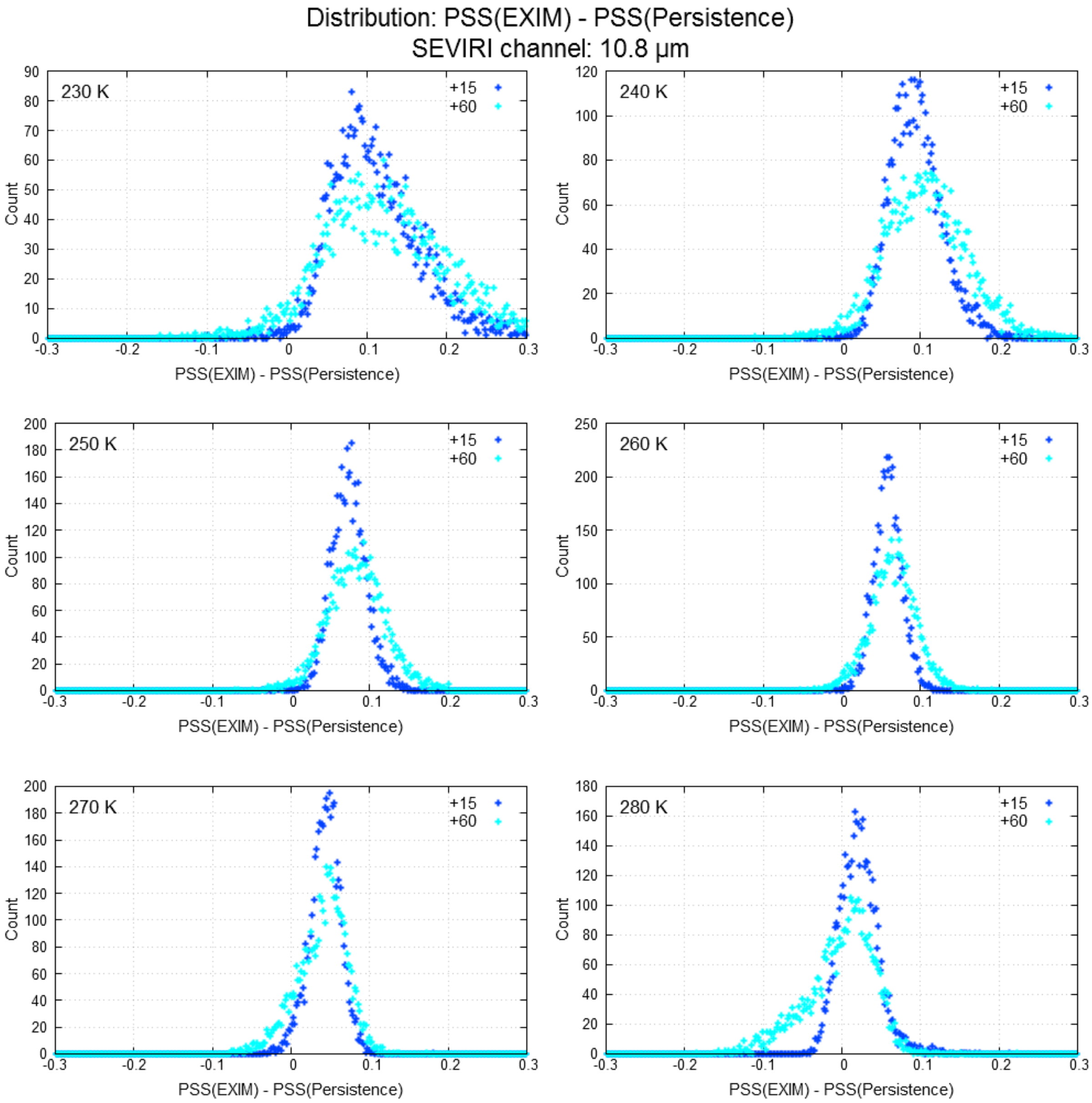

Figure 14: *As in Figure 11 but for SEVIRI channel 10.8µm.*

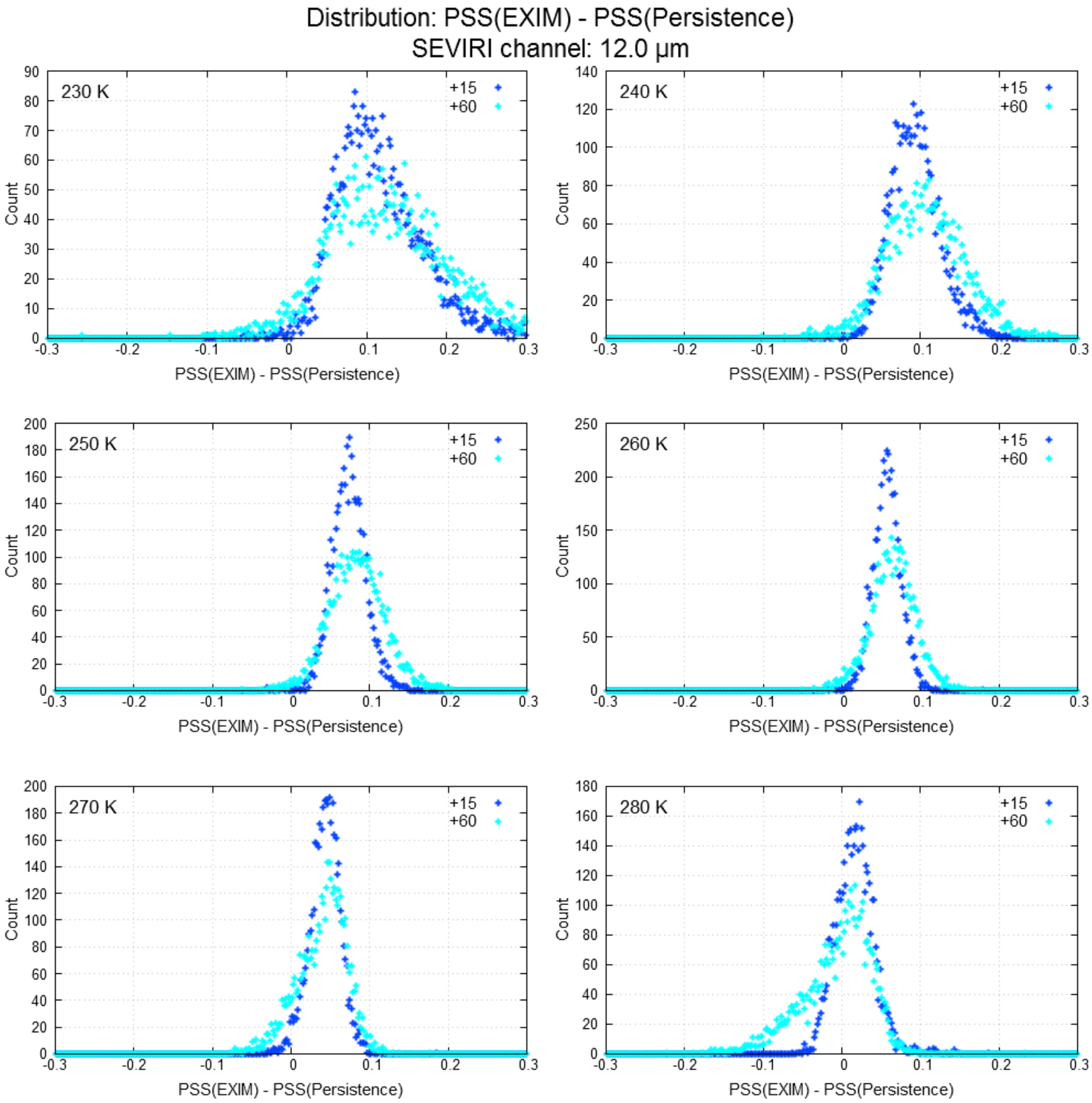

Figure 15: *As in Figure 11 but for SEVIRI channel 12.0µm.*

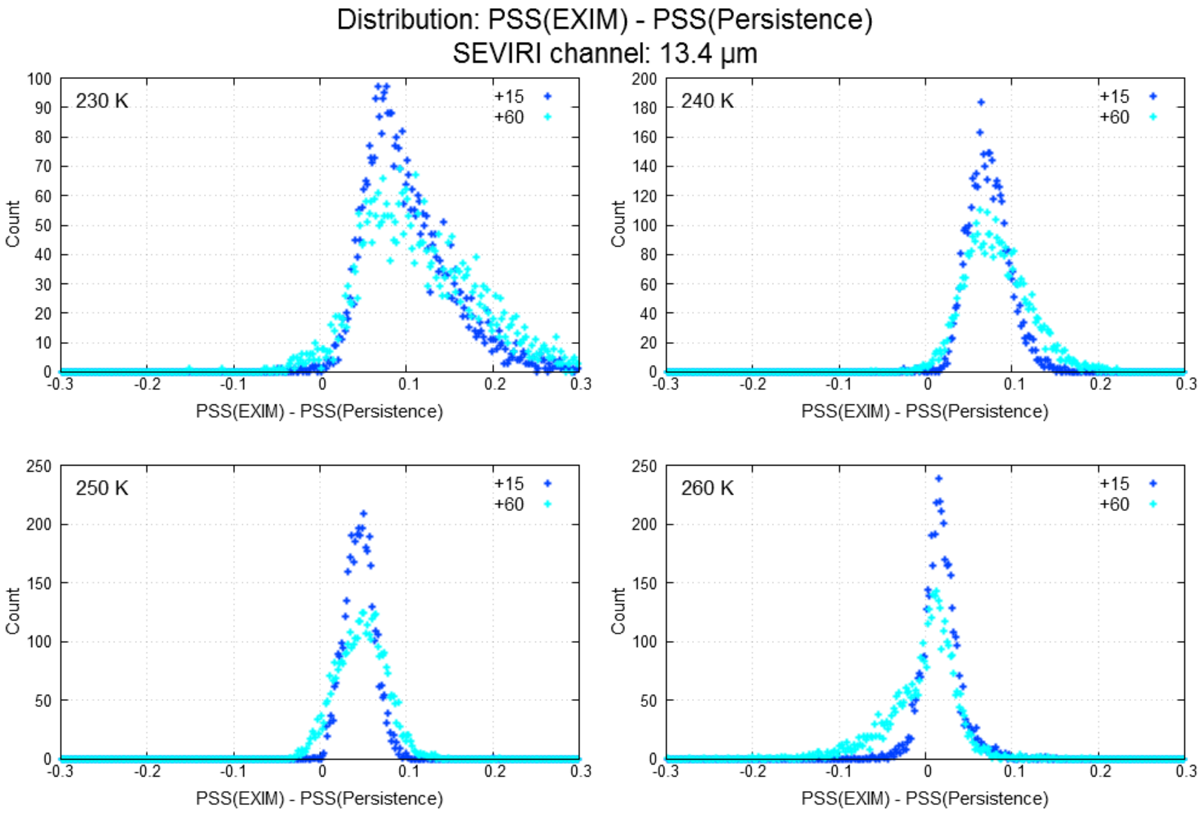

Figure 16: *As in Figure 11 but for SEVIRI channel 13.4µm.*

2.5 THERMAL CHANNELS: WATER VAPOUR

The SEVIRI thermal water vapour channels consist of the spectral bands centred on the wavelengths 6.2 µm and 7.3 µm.

The performance of EXIM compared with persistence is shown in Figure 17 and Figure 18. EXIM outperforms persistence for each threshold, with the poorest performance occurring at the lowest and highest thresholds considered. The reasons for the poor performance of EXIM at the higher thresholds are the same as those for the infrared channels. An explanation for the poor performance of EXIM at lower thresholds may involve the lower values of moisture at higher levels in the atmosphere resulting in imprecise detection.

Plots of the distribution of ΔPSS are shown in Figure 19 and Figure 20. Although EXIM does show an improvement over persistence for all thresholds, the amount of the improvement in ΔPSS is not as great as for the infrared channels.

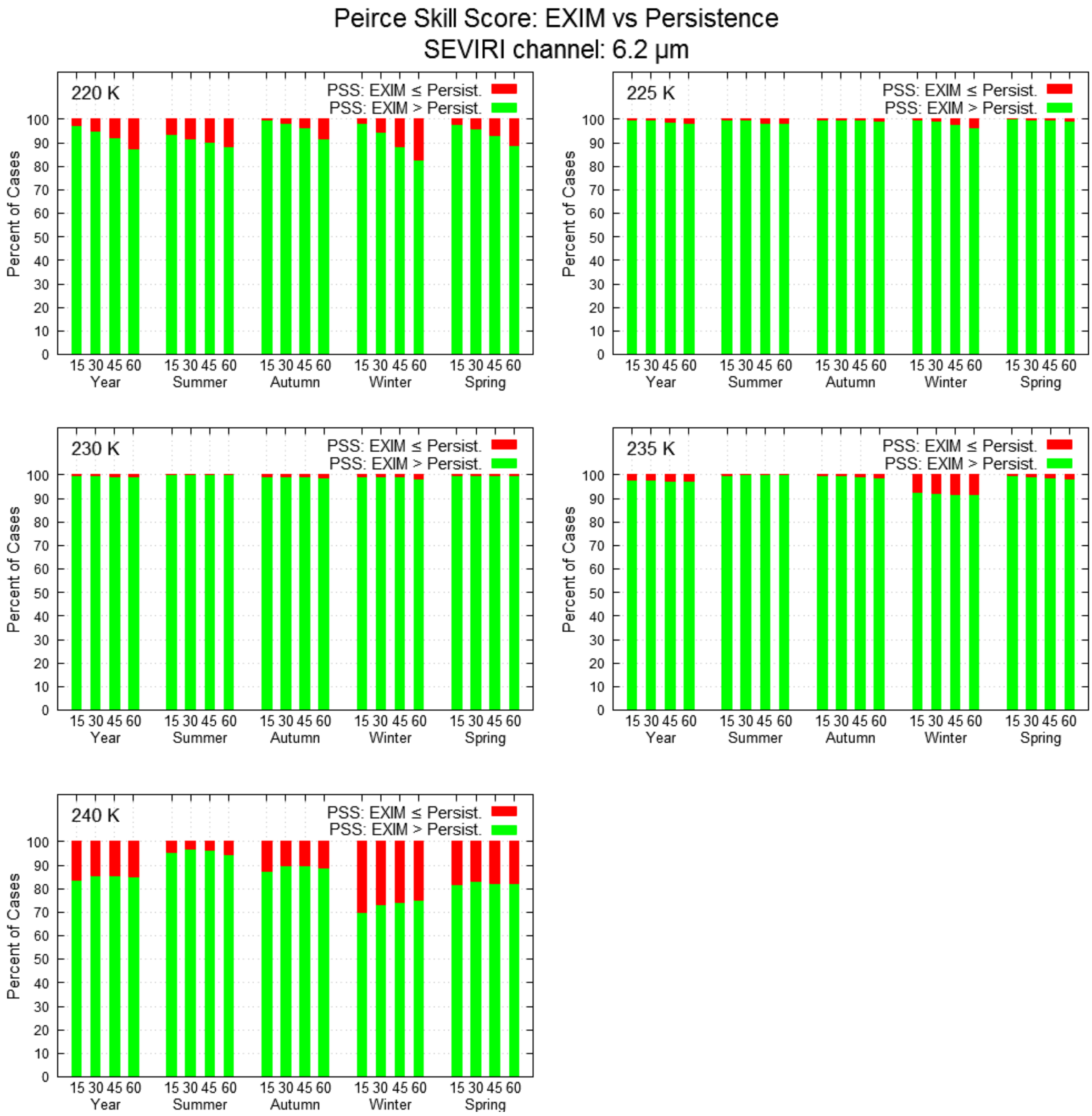

Figure 17: *As in Figure 3 but for SEVIRI channel 6.2µm.*

Figure 18: *As in Figure 3 but for SEVIRI channel 7.3µm.*

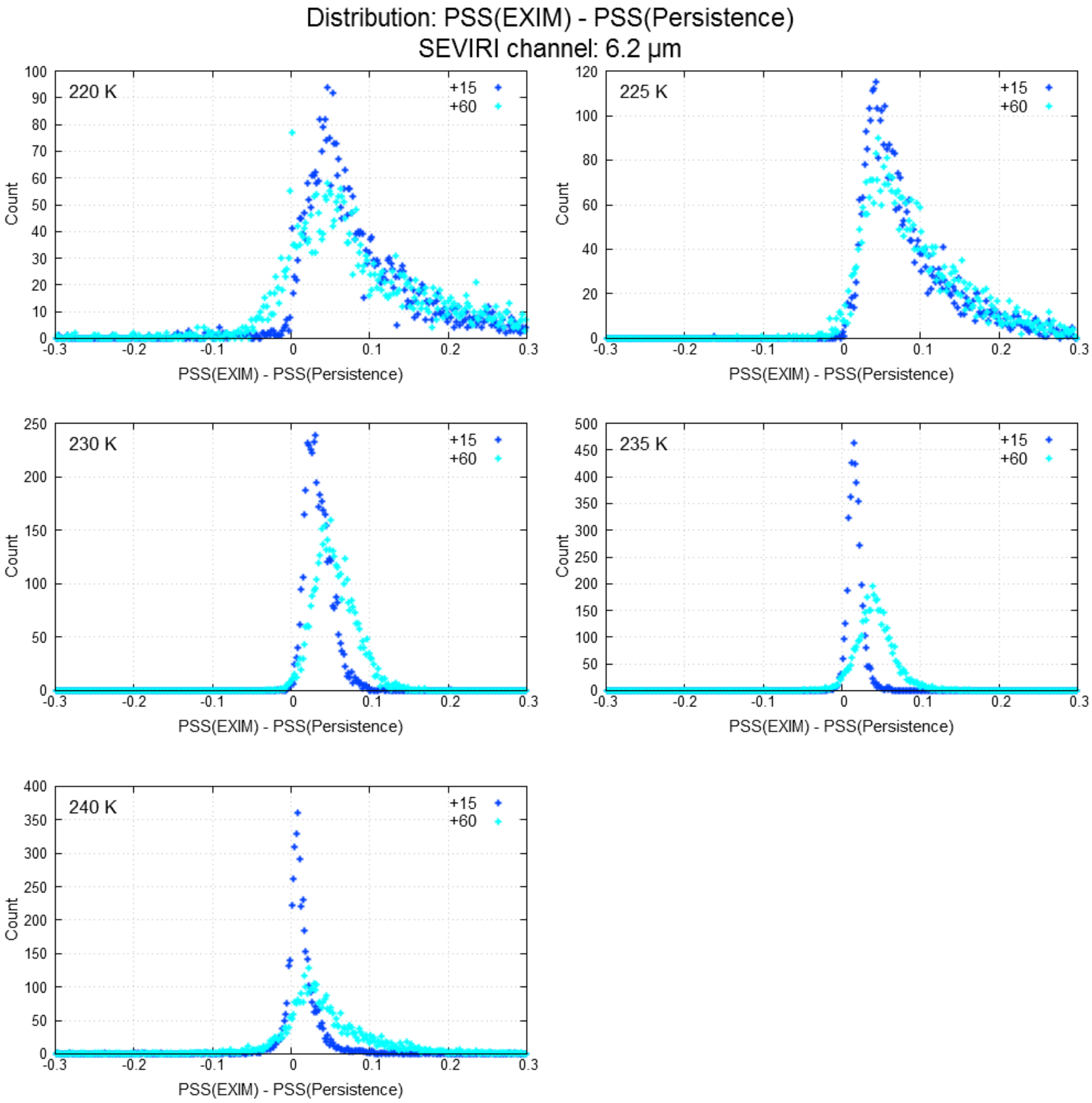

***Figure 19**: As in Figure 11 but for SEVIRI channel 6.2µm.*

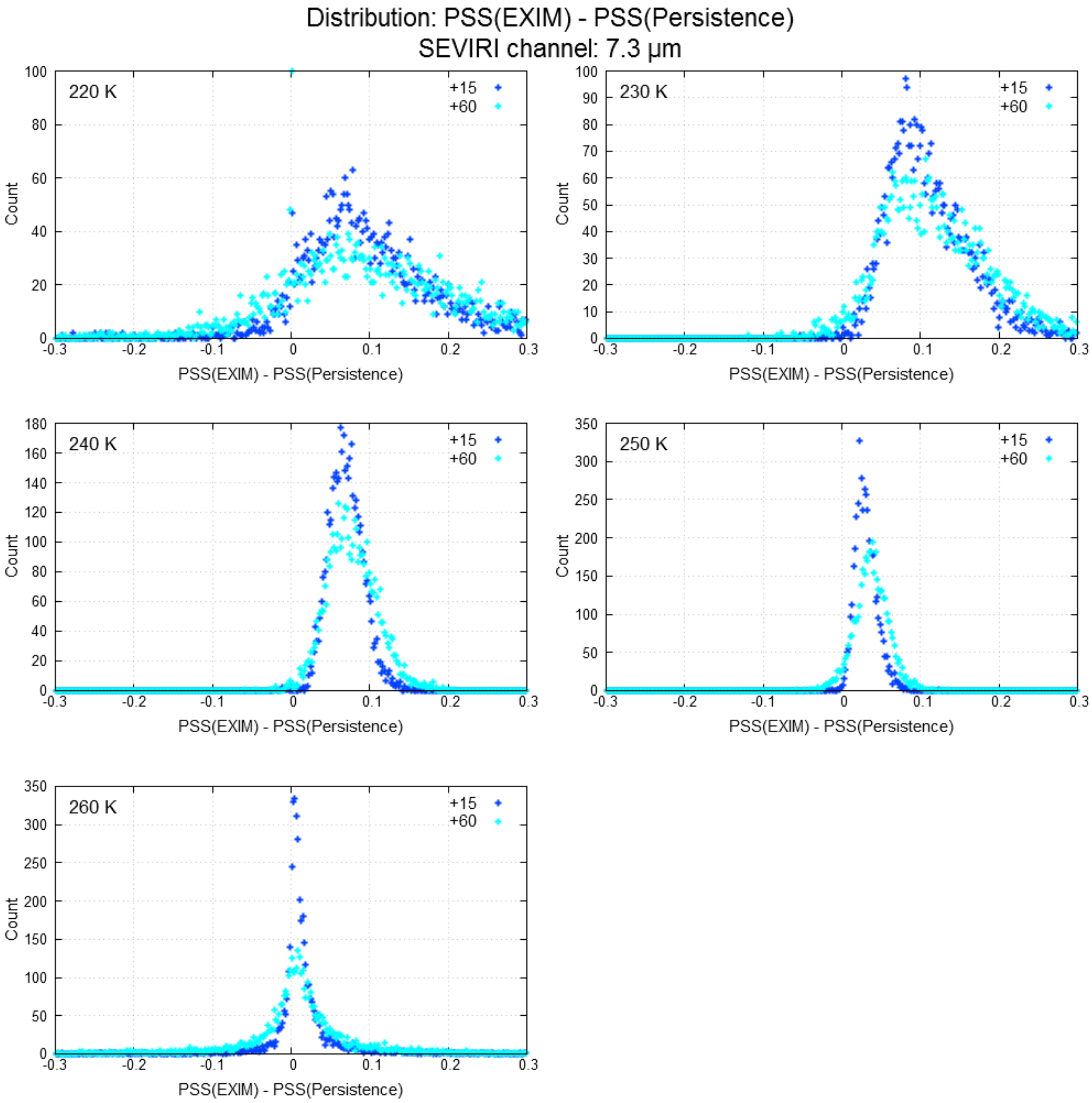

Figure 20: *As in Figure 11 but for SEVIRI channel 7.3µm.*

2.6 VISIBLE CHANNELS

The SEVIRI visible and near-infrared channels consist of the spectral bands centred on the wavelengths 0.6 µm, 0.8 µm, and 1.6 µm.

The results of the visible channels are shown in Figure 21, Figure 22, and Figure 23. Note that there is no winter data available for the validation. The threshold values correspond to radiances, with larger values indicating brighter features, such as clouds, but also snow. Typical dark features include water bodies.

In general, the performance of EXIM for the visible channels is not as good as for the thermal channels. This is because visible channels will detect ground features, which are in general darker than most cloud. This is particularly apparent for the lower threshold values, where the performance of EXIM is poorest. The tracking of ground features also explains why the performance of EXIM relative to persistence decreases with increasing leading time.

Figure 24 shows a time series of ΔPSS = PSS(EXIM) - PSS(Persistence) for the SEVIRI channel 0.8 µm and a threshold value of 8. It is apparent that ΔPSS decreases with increasing lead time, and that the pattern of ΔPSS appears random through the year. That is, there do not appear to be specific weather patterns where EXIM outperforms persistence.

The distribution of ΔPSS plotted for two lead times, +15 minutes (dark blue) and +60 minutes (light blue) for each of the channels and thresholds is presented in Figure 25 – Figure 27. The plots show that EXIM only demonstrates skill for a lead time of +15 minutes.

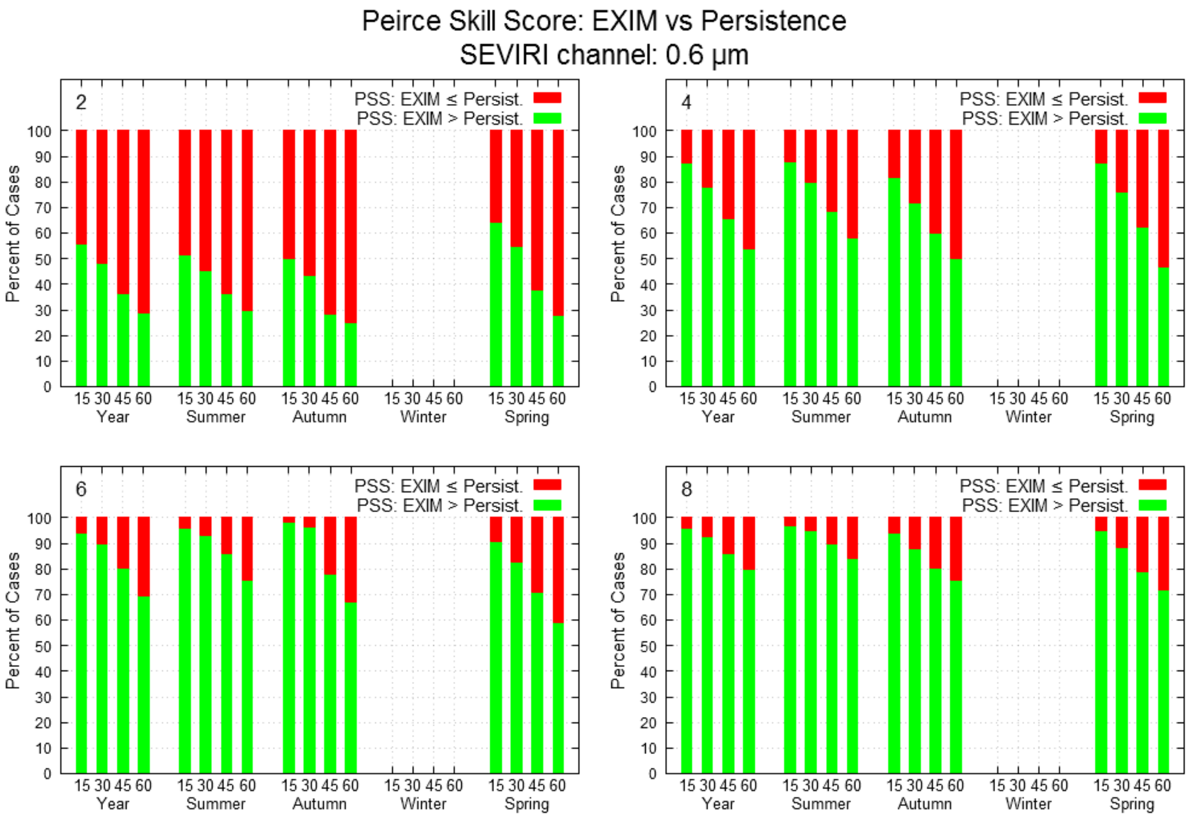

Figure 21: *As in Figure 3 but for SEVIRI channel 0.6µm. The threshold values correspond to radiances, with larger values indicating brighter features.*

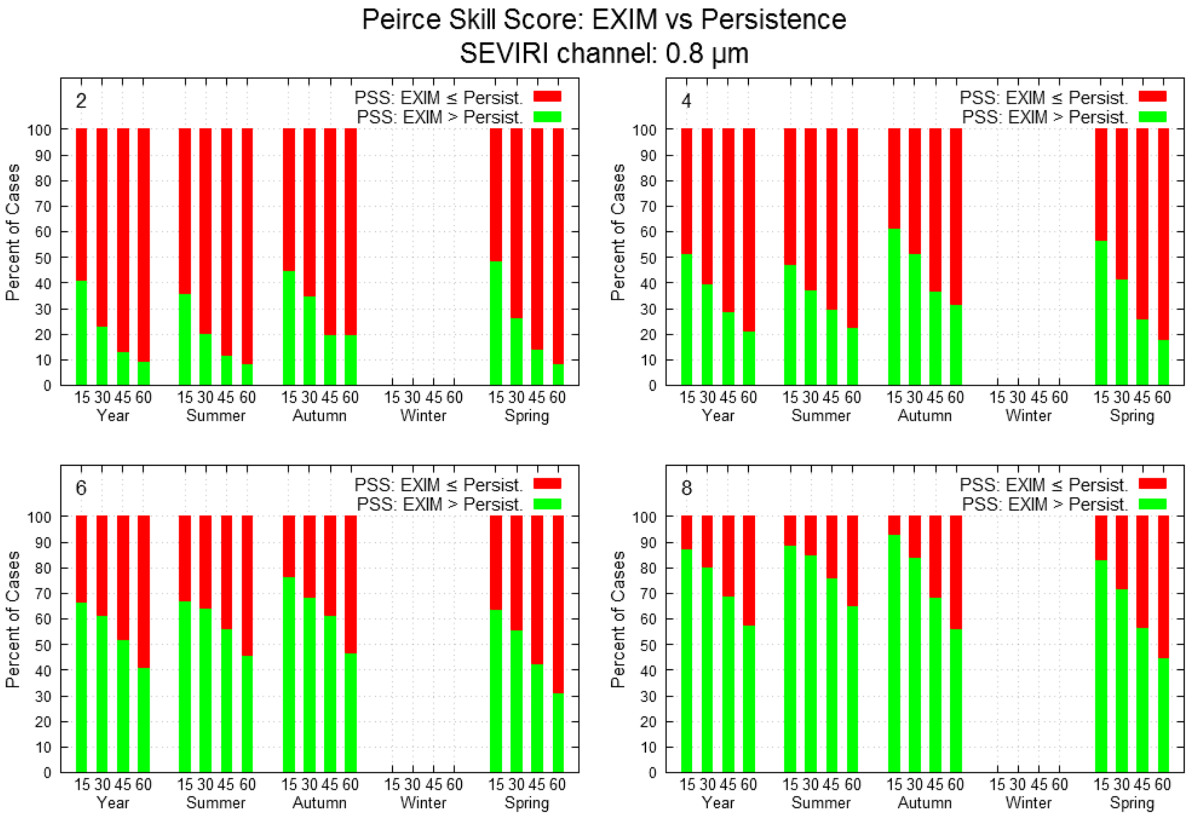

Figure 22: *As in Figure 21 but for SEVIRI channel 0.8µm.*

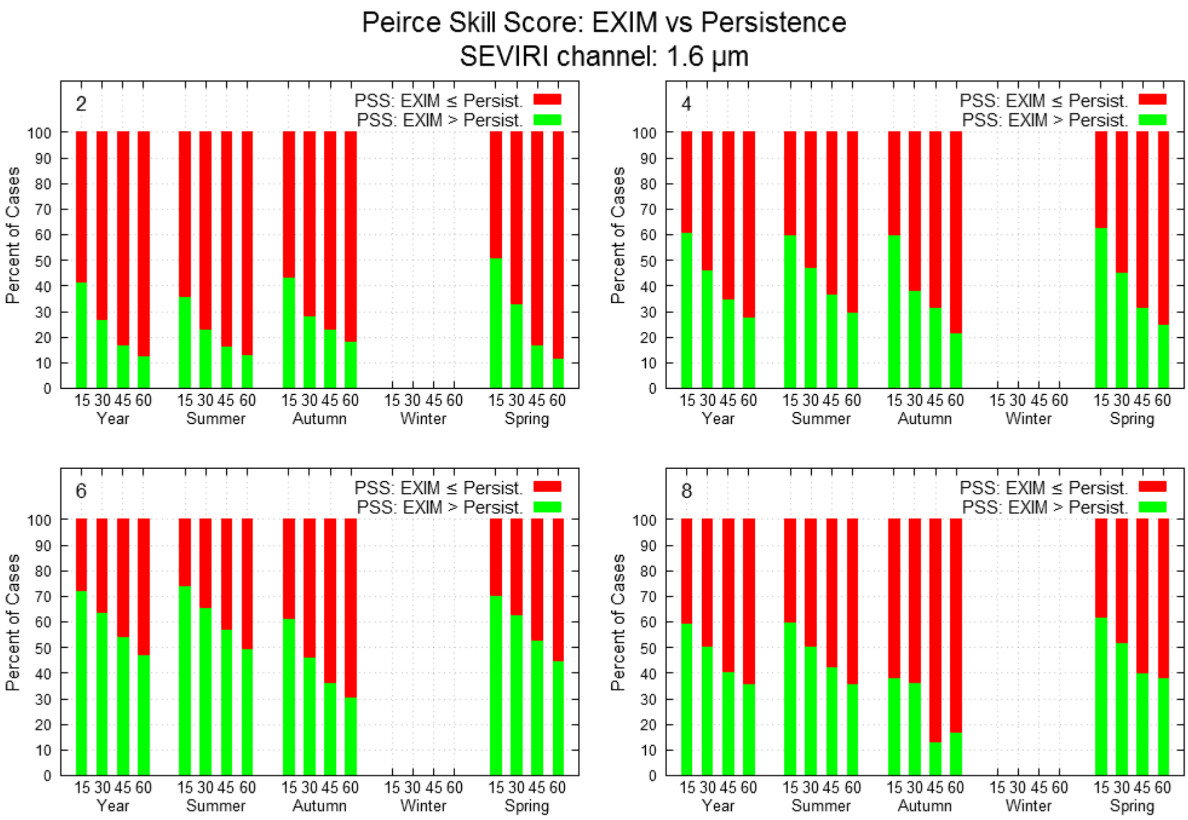

Figure 23: *As in Figure 21 but for SEVIRI channel 1.6µm.*

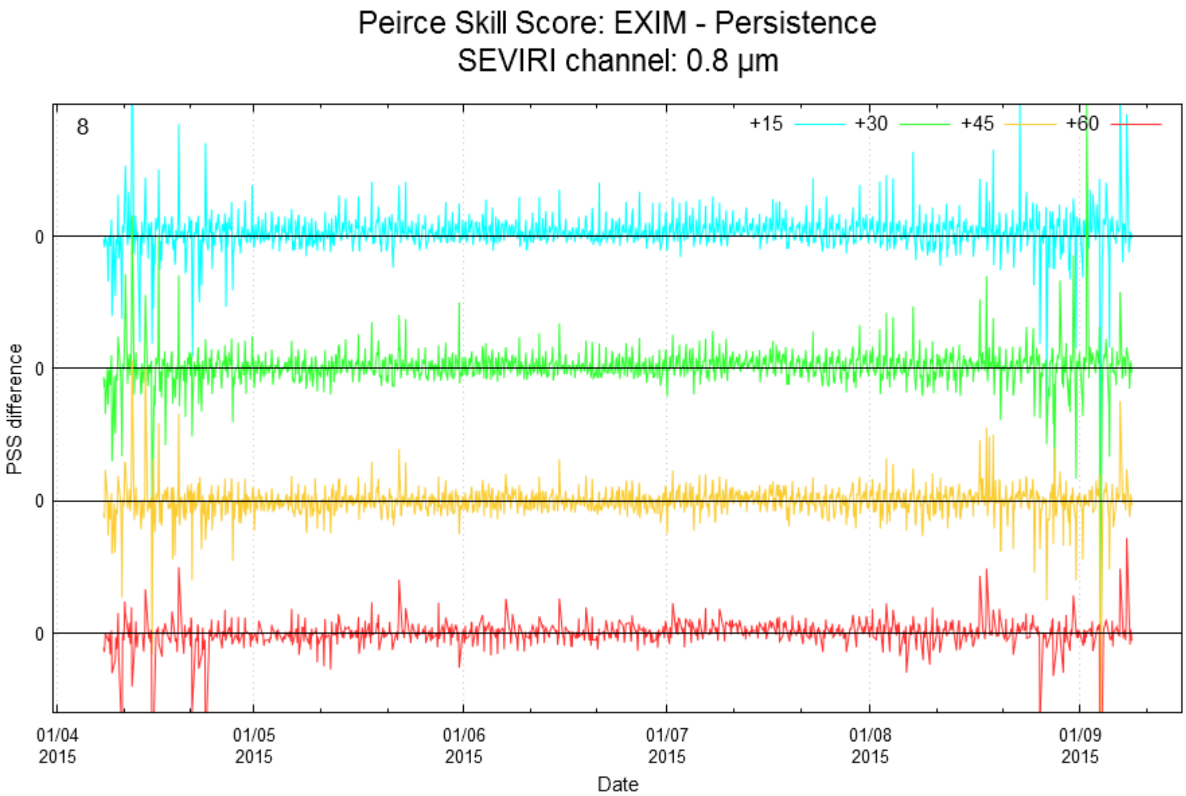

Figure 24: *As in Figure 9 but for SEVIRI channel 0.8 µm and a threshold value of 8.*

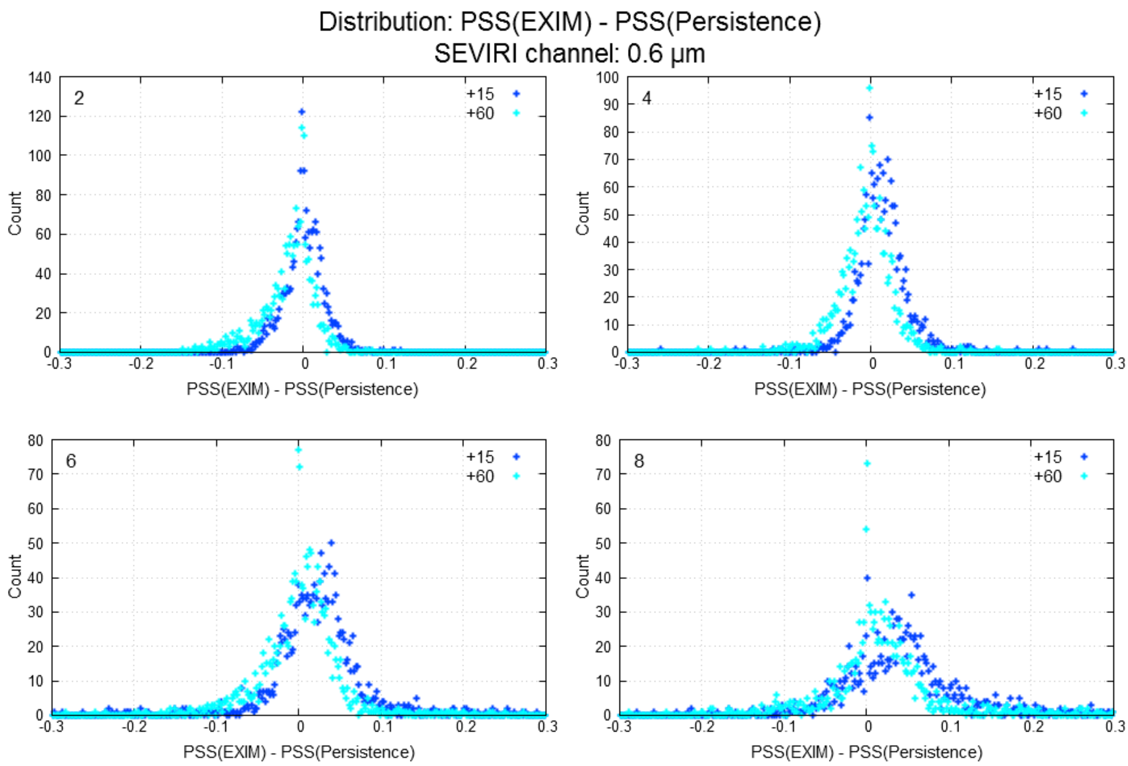

Figure 25: *As in Figure 11 but for SEVIRI channel 0.6µm.*

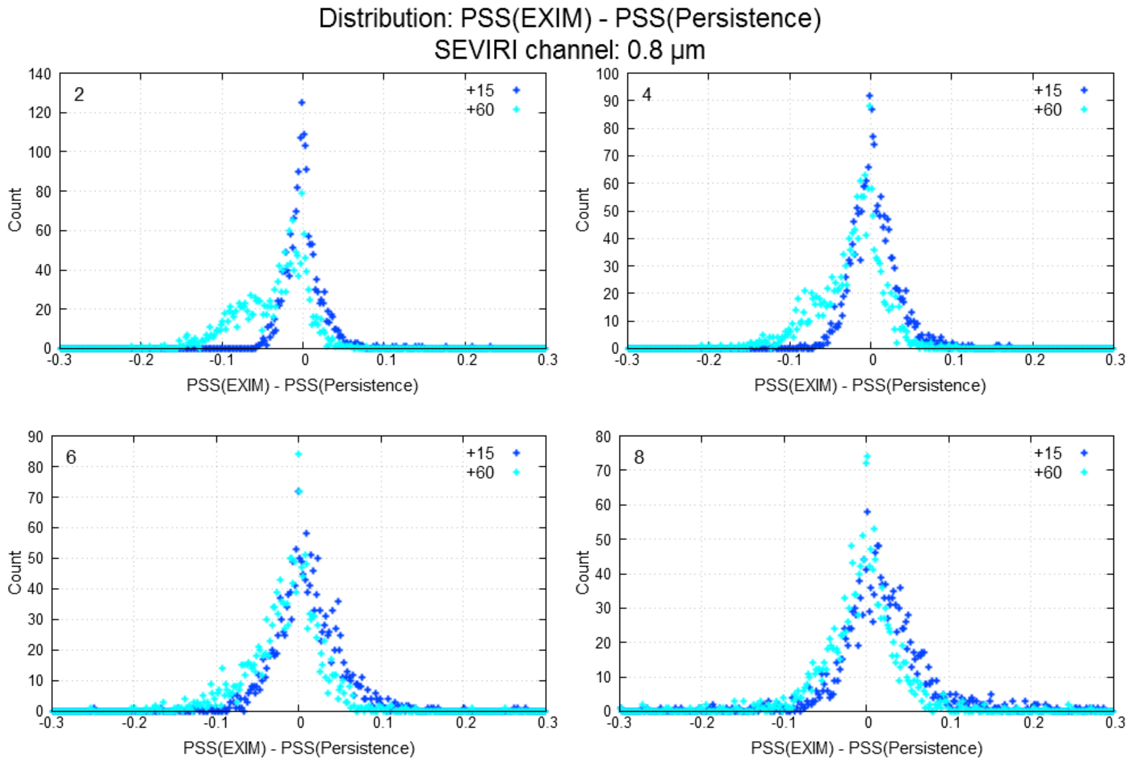

Figure 26: *As in Figure 11 but for SEVIRI channel 0.8µm.*

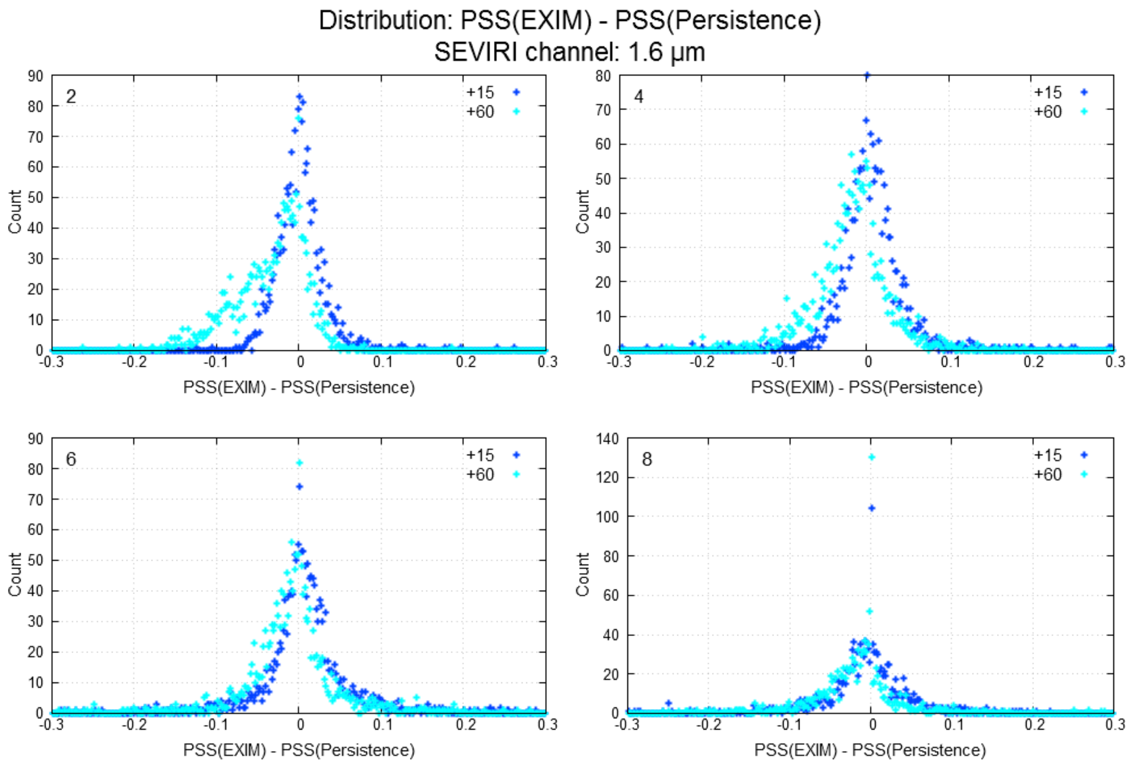

Figure 27: *As in Figure 11 but for SEVIRI channel 1.6µm.*

2.7 CT: Cloud Type

The Cloud Type (CT) product provides a detailed cloud analysis, containing information on the major cloud classes: fractional clouds, semi-transparent clouds, high, medium, and low clouds (including fog) for all the pixels identified as cloudy in a scene. As input it uses at least one of the SEVIRI WV6.2, WV7.3, and IR13.4 channels, along with the IR10.8 and IR 12.0 channels. In addition, the cloud mask product (CMa) and NWP data are used.

The results of the Cloud Type Product are shown in Figure 28. EXIM outperforms persistence for all seasons and almost all lead times (Winter, +60 min is the exception). The performance of EXIM compared with persistence is worse in winter than for the other seasons. This might be attributable to CT's use of the infrared and water vapour channels as input, which, as shown earlier, also exhibited poorer results in winter. During an external review of the document, the hypothesis emerged that the poorer performances may be explained by the different seasonal cloud types. In winter, low cloud and fog tend to be more prevalent than in other seasons. Both cloud types are characterized by a uniform appearance in the visible and infrared spectral channels. The lack of suitable tracers (cloud features) within these uniform layers means that the AMVs are sparsely distributed, typically concentrating only on the boundaries of the cloud region. AMVs at the cloud or fog boundary will be assumed by EXIM to apply to a neighbouring region centred on the AMV, and will result in the entire cloud patch translating, when in fact, it could just be a local movement of the boundary, due to say subsidence drying. In these cases Persistence will perform better than EXIM. It is also true that EXIM in its current version makes no attempt to distinguish displacements at different atmospheric levels, so lower-level features may be displaced by high-level vectors. This suggests as a first step to attempt extrapolation with vertically differential displacement fields, and continue investigating on the actual root cause of the found PSS patterns in case they persist.

Figure 29 shows the distribution of ΔPSS for the lead times +15 min and +60 min. At a lead time of +15 min EXIM has a PSS approximately 0.06 better than persistence, which drops to approximately 0.02 for a lead time of +60 min.

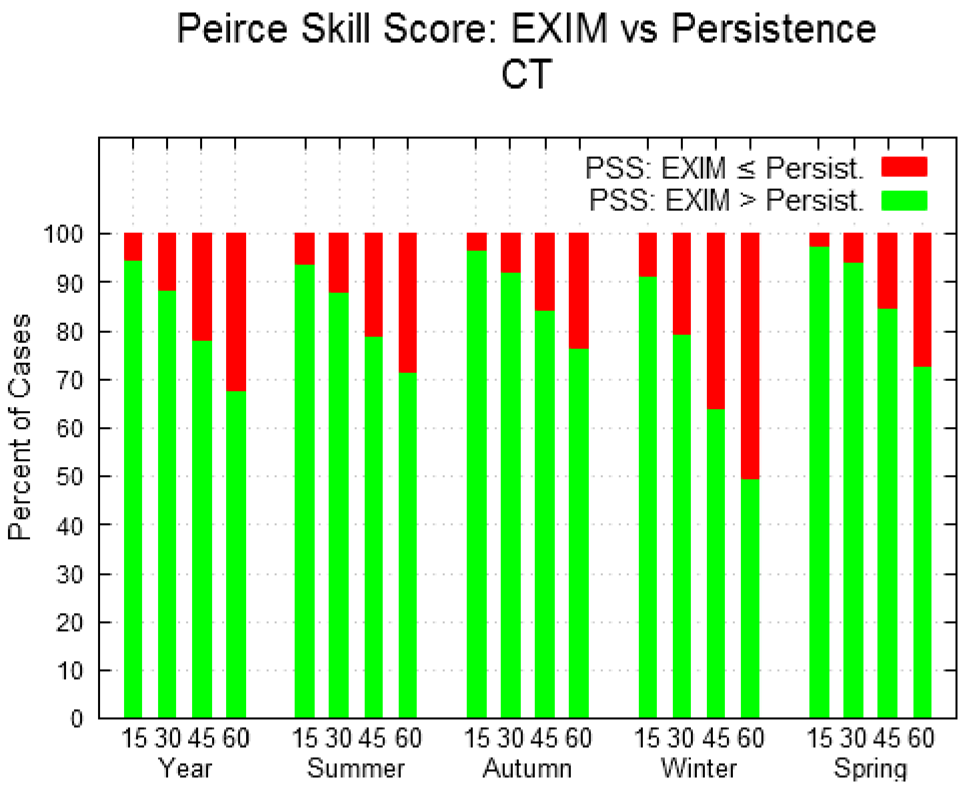

Figure 28: *As in Figure 3 but for Cloud Type.*

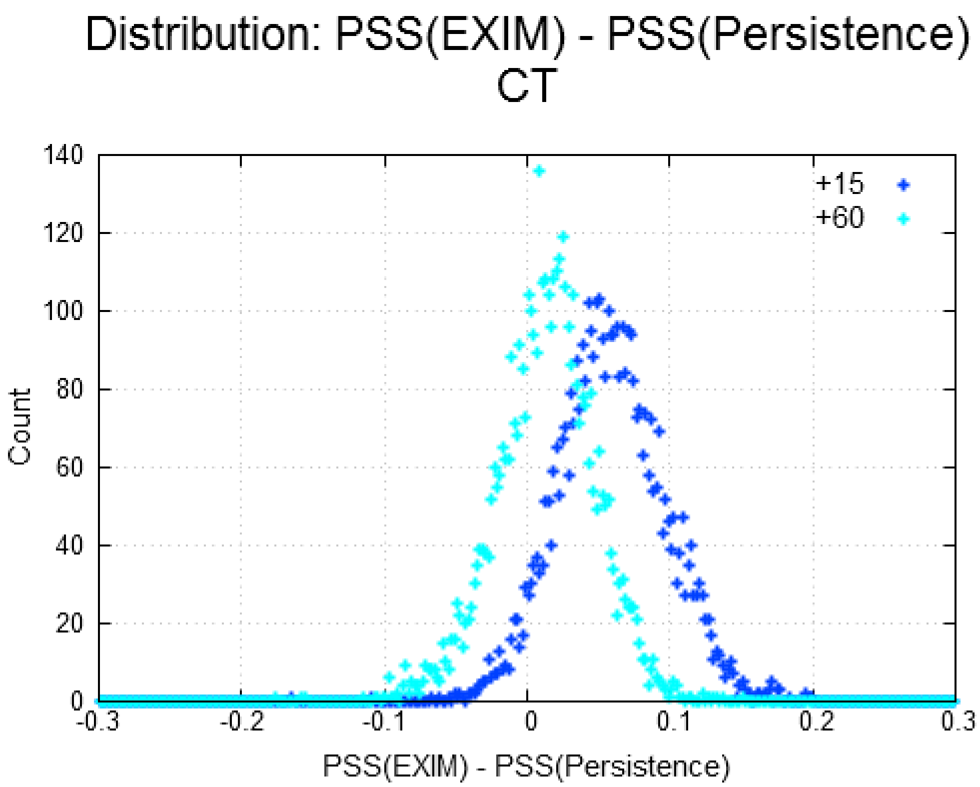

Figure 29: *As in Figure 11 but for Cloud Type.*

2.8 CTTH: Cloud Top Temperature and Height

The Cloud Top Temperature and Height (CTTH) product is used for the analysis and early detection of thunderstorm development, providing information on the cloud top temperature and height for all pixels identified as cloudy in the satellite scene. As input it uses the SEVIRI channels IR10.8, IR12.0, and at least one of WV6.2, WV7.3, or IR13.4. In addition, it uses input from CMa, CT, and other data.

The results of CTTH are shown in Figure 30, with the thresholds showing cloud top heights in metres. EXIM outperforms persistence for low to mid-level clouds (≤ 8000 m) for all seasons (except winter at 8000 m) and lead times. Of note is that EXIM performs marginally poorer in winter..

The performance of EXIM for clouds at high levels (≥ 8000m) becomes poorer relative to persistence, which contrasts its good performance for high clouds in the thermal channels (Section 2.4). EXIM exhibits a decreasing trend with increasing lead time in performance for low cloud top heights (≤ 6000 m). However, for thresholds of 7000 m and greater, there are cases where the performance of EXIM improves with increasing lead time (e.g. threshold 7000, Autumn, Winter, Spring, and subsequently all seasons for greater thresholds). Of course, one could again speculate (c.f. Section 2.7) that vertically differential displacement fields would help in improving the EXIM skill scores. However, there are also specifics of the CTTH product that are detrimental to the scores (and need to be suppressed in future validations of EXIM-CTTH): A graphical examination of the CTTH product reveals that at levels at and above about 7000 m, the CTTH field exhibits coarse blocks consisting of 16×16 pixels, rather than the typical fine pixel scale structure present at the lower heights. An example of this pattern can be seen in Figure 31. These coarse blocks are a regular feature that can be explained from the CTTH derivation where several tests are carried out only segment-wise (configurable; default size being used here: 16×16 pixels; cf. [RD.3]) in order to reduce computation time. As the fine scale clouds move from one time step to the next, these coarse blocks remain stationary until the cloud shifts to the next 16×16 pixel block. That is to say, a number of time steps are required before the blocky cloud feature moves to the neighbouring block. This delayed movement in the coarse blocks is the reason why Persistence performs better for shorter lead times, whereas EXIM performs comparatively better for later lead times.

Figure 32 shows the distribution of ΔPSS for the lead times +15 min and +60 min. The peak in the distribution for both lead times is located at PSS > 0 for cloud top heights ≤ 9000 m, although the spread increases progressively with height above 6000 m. As indicated earlier in Figure 30, Figure 32 also shows that at heights of 6000 m and above EXIM produces better results at a lead time of +60 min than at +15 min. The improvement in performance that EXIM brings is, on average, at most 0.05.

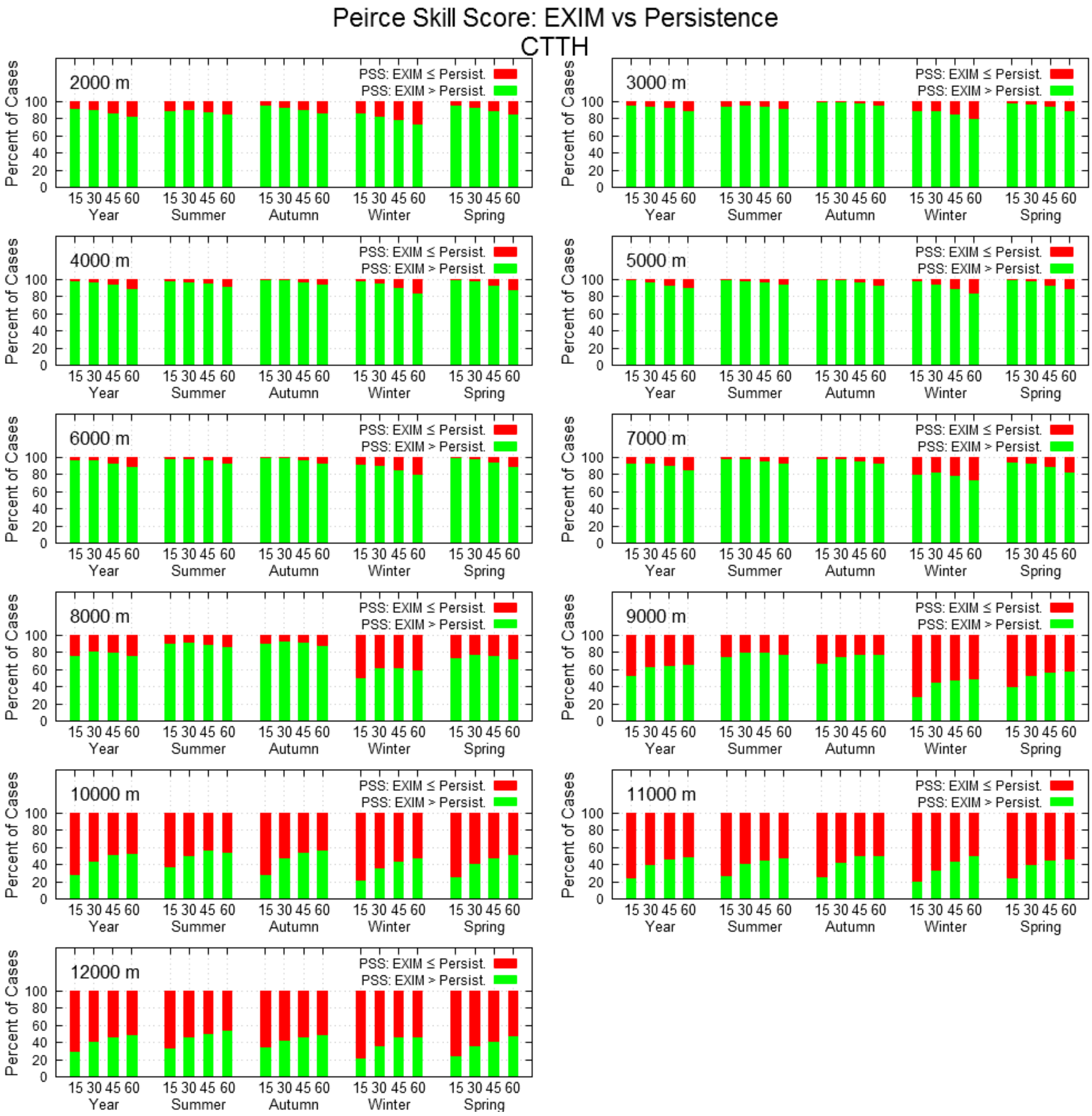

Figure 30: *As in Figure 3 but for the Cloud Top Temperature and Height product.*

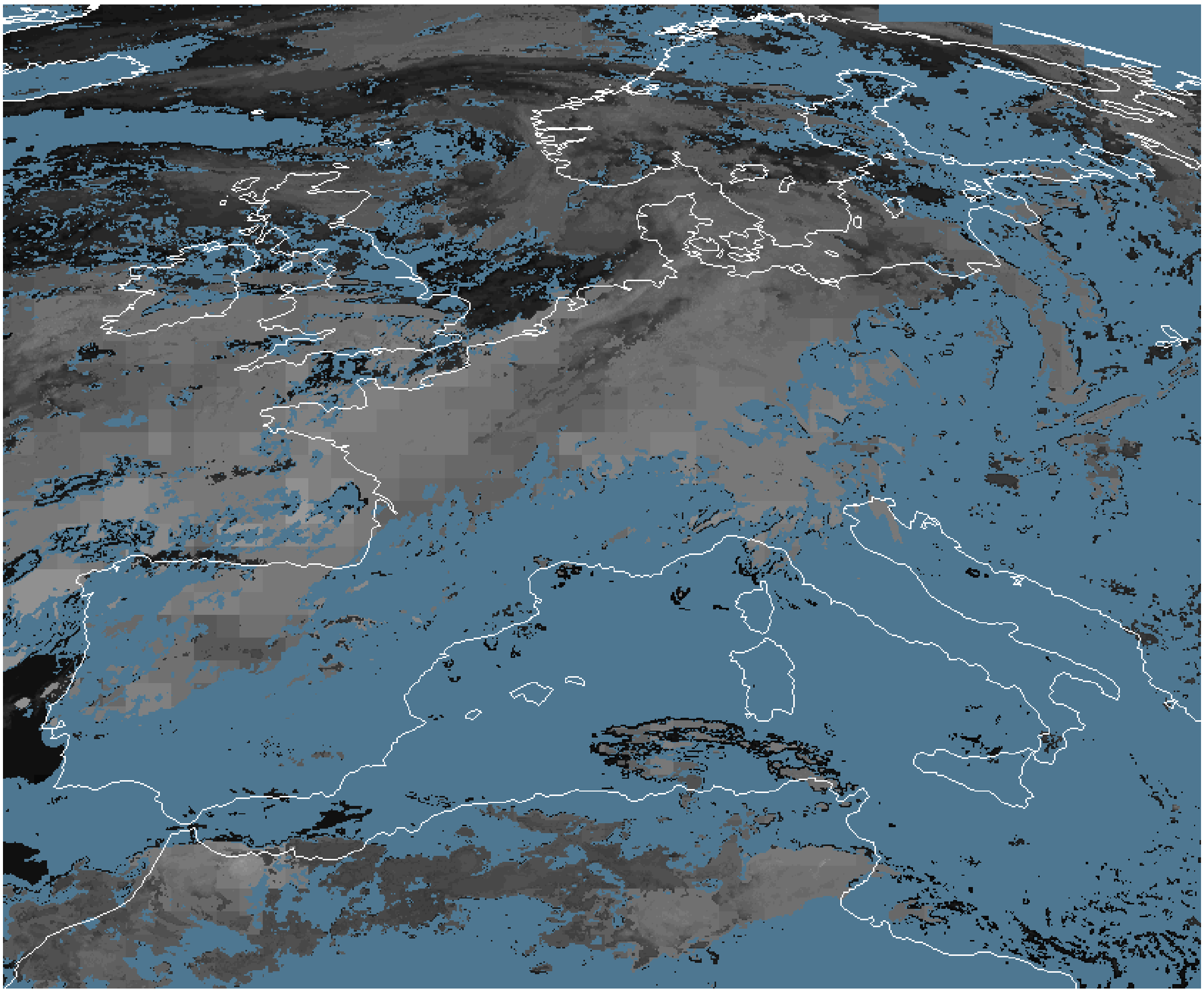

Figure 31: *CTTH product on 2 September 2016, 2100 UTC; the blocky structure referred to in the text can be seen over central and western Europe.*

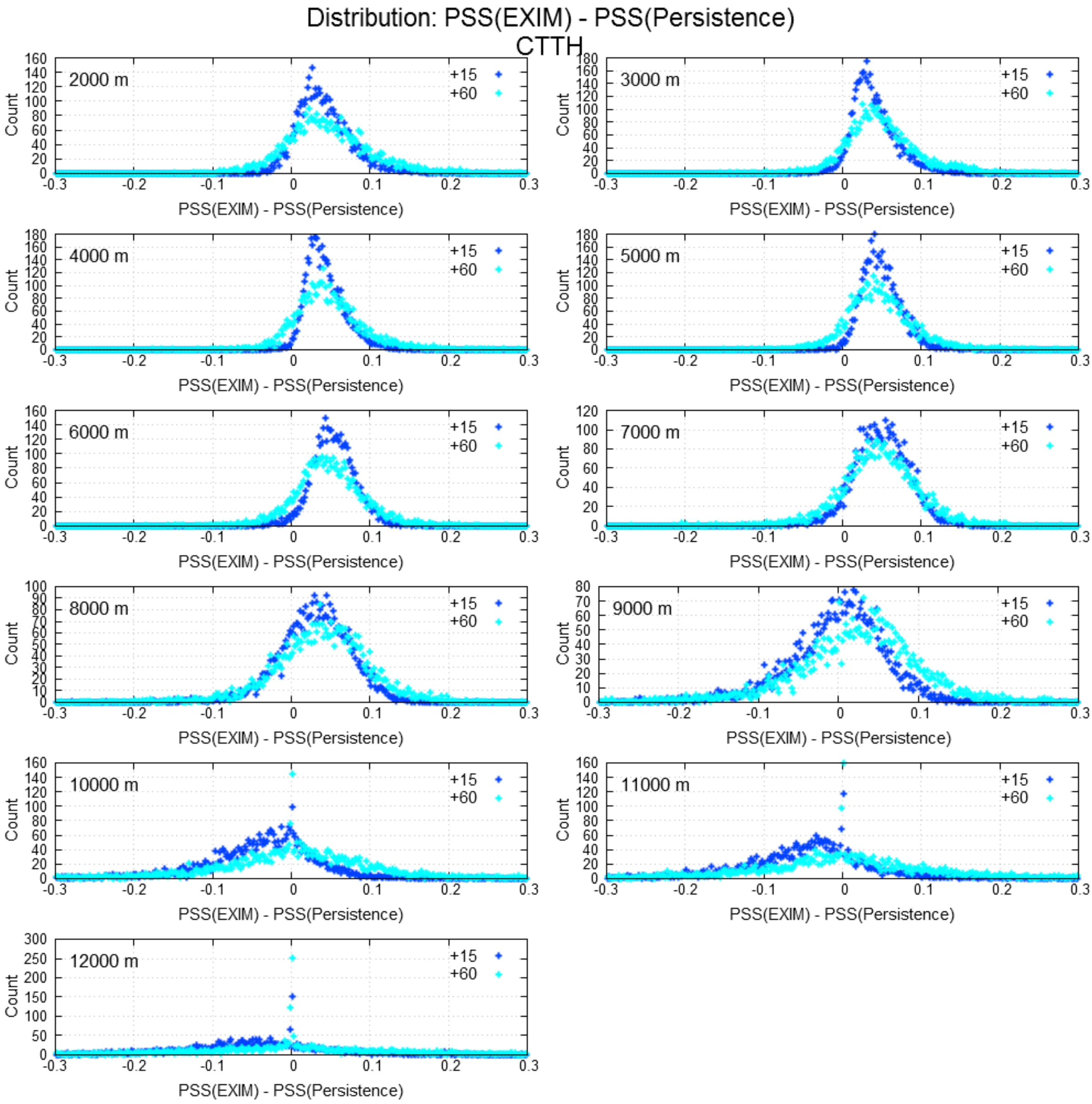

Figure 32: *As in Figure 11 but for Cloud Top Temperature and Height.*

2.9 CRR: CONVECTIVE RAINFALL RATE

The Convective Rainfall Rate (CRR) product estimates the rain rate from convective clouds. Along with other input products, CRR uses as input the SEVIRI IR10.8 and WV6.2 channels and optionally the VIS0.6 channel.

The results of the EXIM forecasts of the Convective Rainfall Rate (CRR) are presented in Figure 33. The threshold represents a class in rain rate, shown in Table 4, with lower thresholds representing lower rain rates. As higher thresholds occur less frequently, the number of cases used in the analysis needs to be taken into account (see Figure 59) when interpreting the percentages in Figure 33. When the number of cases is low, the effect of randomness dominates in the results (e.g., Figure 33 threshold 10, spring, where only 1 case was observed).

CRR Class Threshold	Rain rate (mm/h)
1	[0.2, 1)
2	[1, 2)
3	[2, 3)
4	[3, 5)
5	[5, 7)
6	[7, 10)
7	[10, 15)
8	[15, 20)
9	[20, 30)

Table 4: *CRR class and corresponding rain rate.*

In general, EXIM outperforms persistence for lower thresholds (≤ 4), with winter once again being the season where EXIM performs worse. For the middle thresholds (4, 5) EXIM performs best in autumn, and in spring for threshold 6. At higher thresholds EXIM performs poorer than persistence.

One interesting feature about the CRR product is that the greatest number of rain events were detected in autumn regardless of rain rate (Figure 59). This is particularly noteworthy for the higher rain rates, which may be expected to occur more frequently in summer. A possible explanation for this anomaly could be that the 3 mm/h rain rate threshold that CRR uses to distinguish convective and stratiform rain types, and to estimate convective rainfall rate, is too low. Using a higher threshold (e.g. 5 mm/h) may decrease the bias of cases detected by the CRR algorithm in autumn.

Figure 34 shows a time series of ΔPSS = PSS(EXIM) - PSS(Persistence) for CRR and a threshold value of 2. In general ΔPSS decreases with increasing lead time, and there are periods where EXIM performs better than persistence and vice versa. For example, during the period 1-6 May EXIM outperformed persistence, as a large part of Europe was dominated a low pressure system centred west of the British Isles. This created a southwesterly flow over mainland Europe leading to fast moving rain areas, with hardly any regions of stationary precipitation.

Conversely, in the period 1-3 April persistence outperformed EXIM, as the flow was from the northwest which produced blocking precipitation on the northern side of the Alps. Additionally, the presence of low pressure systems over the Mediterranean seems to result in a poorer performance of EXIM, presumably due to the near-stationary precipitation regions near the centre of the low pressure system.

The CRR product has the option to use the VIS0.6 channel. As EXIM produces relatively poor results for this channel, the results of EXIM applied to CRR only at night (1800 – 0600 UTC) were examined (not shown). EXIM gave better results for thresholds ≤ 6 for all seasons. At higher thresholds, improvements in EXIM were apparent for some lead times and seasons, but not for all.

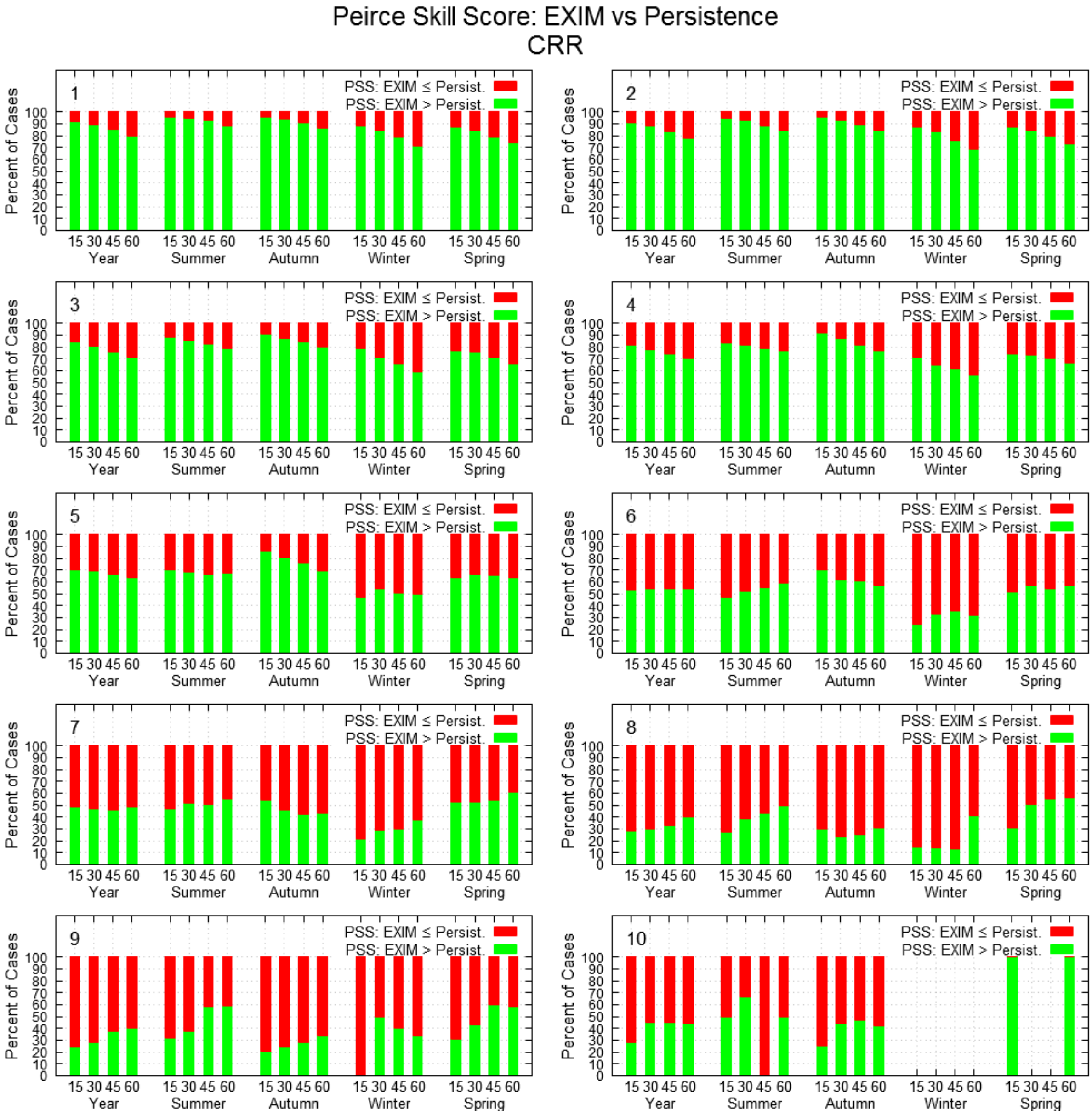

Figure 33: *As in Figure 3 but for Convective Rainfall Rate Product. The thresholds correspond to the rain rates shown in Table 4.*

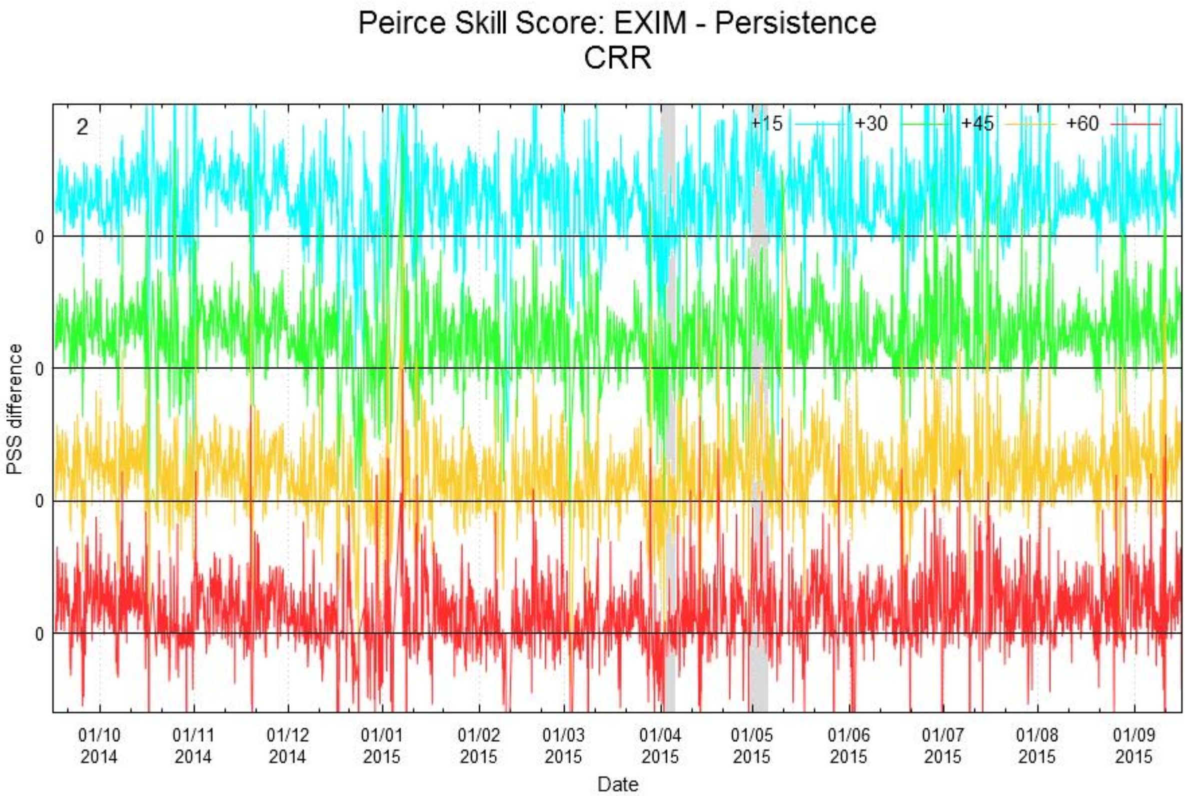

Figure 34: *As in Figure 9 but for CRR and a threshold value of 2. The two grey columns mark the dates mentioned in the text.*

Figure 35 shows the distribution of ΔPSS for the lead times +15 min and +60 min. At lower thresholds (≤ 5) EXIM is on average 0.1 better than persistence at +15 min lead time. The PSS drops for later lead times, with EXIM being marginally better than persistence after +60 min for the low thresholds. At higher thresholds, the limited number of cases prevents a detailed analysis, but the indication is that EXIM performs poorer than persistence.

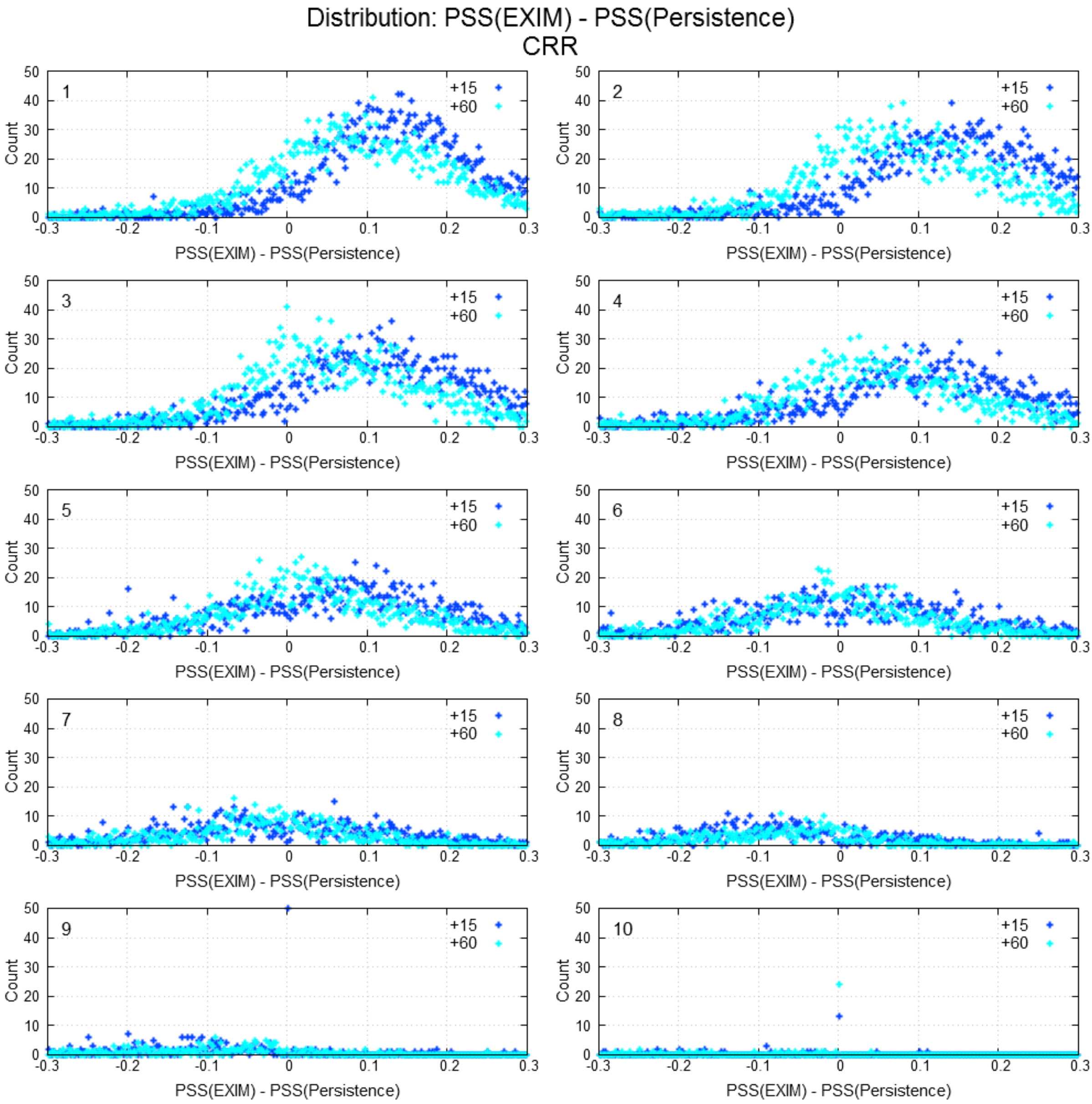

***Figure 35:** As in Figure 11 but for Convective Rainfall Rate.*

2.10 CRPh: Convective Rainfall Rate from Cloud Physical Properties

The Convective Rainfall Rate from Cloud Physical Properties (CRPh) product estimates rain rates from convective clouds through cloud top microphysical information such as the cloud top effective radius, and cloud optical thickness. Along with other input products, CRPh uses the SEVIRI IR10.8 channel.

The results of the forecasts of the CRPh are presented in Figure 36. The threshold values correspond to rainfall rates scaled by the factor 0.2 mm/h (i.e. 1 = 0.2 mm/h, 11 = 2.2 mm/h, 21 = 4.2 mm/h, etc.). The CRPh product is known to degrade in quality for poor illumination conditions and the quality flags currently do not allow a proper masking of less illuminated pixels. Thus only the 1200 UTC data were used as it would be certain it would be usable over the entire domain.

As for the CRR product, EXIM outperforms persistence for lower thresholds (≤ 41) at all lead times. There are some anomalous results, whereby the performance of EXIM is better for a lead time of 30 minutes than for 15 minutes (e.g. threshold 1 Summer, threshold 31 Spring, etc**)**. At higher thresholds, EXIM has skill only for the shortest lead times and only in summer and autumn. For the highest thresholds, the total number of cases need to be taken into account (see Figure 60), since higher thresholds are observed less often, and the effect of randomness plays a larger role in the results. One interesting feature of Figure 60 is that the greatest number of events occurs in winter/autumn for all thresholds. This may be reasonable for low rain rates, but one would expect that high rain rates would occur preferentially in the warmer months (e.g. spring, summer). This differs from the CRR product which produced the most precipitation events in autumn regardless of intensity.

The thresholds in panels Figure 33 and Figure 36 refer to different rain rates, and so the panels cannot be directly compared with each other. The panels can be broadly grouped into low, medium, and high rainfall intensity, as shown in Table 5. The comparison shows that EXIM generally performs equally well with CRR and CRPh.

Rainfall intensity	CRPh Thresholds	CRR Class Thresholds
Low	1,11,21	1,2,3,4
Medium	31,41	5,6
High	51,61,71,81,91	7,8

Table 5: *Comparison of the rain rate thresholds between Figure 33 and Figure 36.*

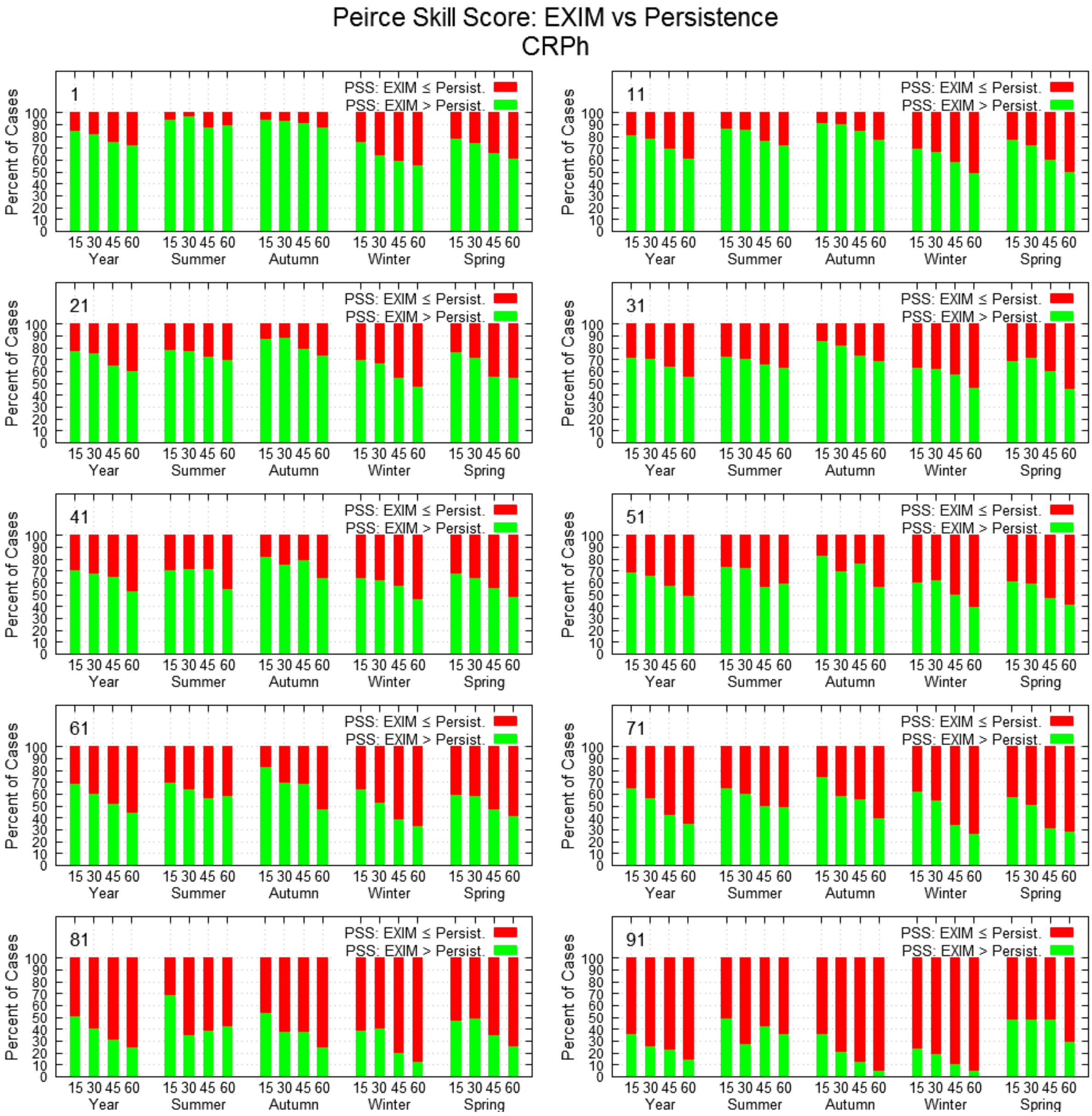

Figure 36: *As in Figure 3 but for Convective Rainfall Rate from Cloud Physical Parameters Product. The threshold values correspond to rainfall rate scaled by the factor 0.2 mm/h (i.e. 1 = 0.2 mm/h, 11 = 2.2 mm/h, 21 = 4.2 mm/h, etc.).*

2.11 PC: PRECIPITATING CLOUDS

The Precipitation Clouds (PC) product gives the likelihood of precipitation. As input it uses the SEVIRI channels VIS0.6, NIR1.6, IR3.9, WV6.2, WV7.3, IR10.8, and IR12.0 during the day and the channels WV6.2, WV7.3, IR10.8, and IR12.0 at night. In addition, it uses input from CT and other data.

The results of the forecasts are presented in Figure 37. Only the data during the day (1200 UTC) has been used for the analysis, in order to avoid PC's night algorithm, which is known to yield considerably different (and worse) results. The threshold values correspond to precipitation likelihood with higher values corresponding to higher precipitation probabilities. The likelihood as a percentage can be obtained by multiplying the threshold value by 10. In general, EXIM performs better than persistence for lower precipitation probabilities (≤ 3), but worse than persistence for higher precipitation probabilities.

The performance of EXIM is poorest in winter. This is due to PC using the infrared channels as input, and these were shown in Sections 2.4 and 2.5 to produce poorer results in winter.

Due to the small number of cases analysed (see Figure 61), the distribution plots of ΔPSS did not yield useful data and are thus not shown.

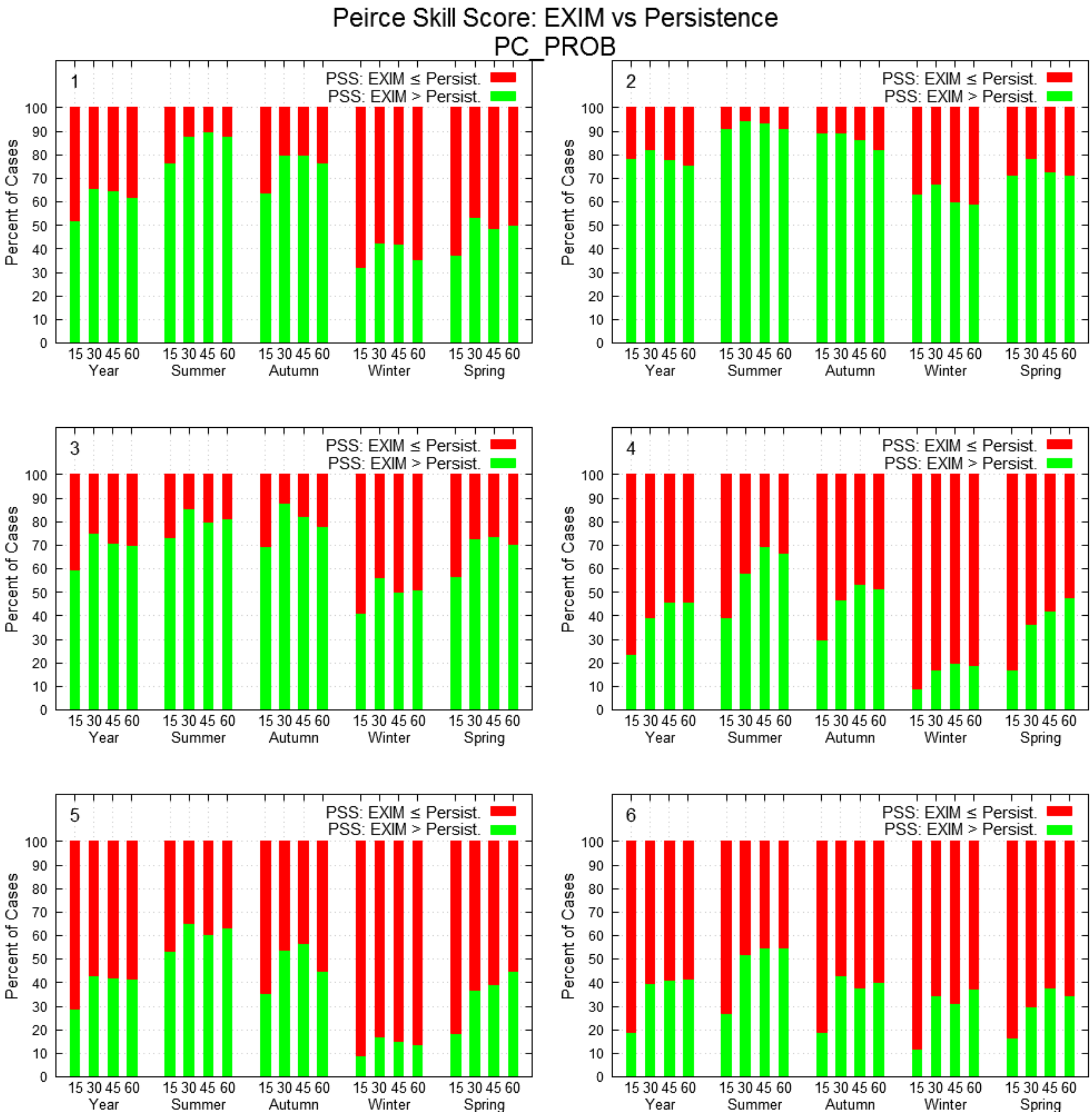

Figure 37: *As in Figure 3 but for Precipitating Clouds Product. The threshold values correspond to precipitation likelihood with higher values corresponding to higher precipitation probabilities.*

2.12 PCPh: Precipitating Clouds from Cloud Physical Properties

The Precipitation Clouds from Cloud Physical Properties (PCPh) product provides an estimation on the probability of precipitation occurrence which is similar to PC, but instead uses the cloud top microphysical properties, effective radius, and cloud optical thickness. As input it uses, amongst other products, CT and the SEVIRI IR10.8 channel.

The results of the forecasts of PCPh are presented in Figure 38. As for PC, only the data during the day (1200 UTC) have been used for the analysis. The threshold values correspond to precipitation likelihood with higher values corresponding to higher precipitation probabilities. In general, EXIM performs better than persistence for all seasons and lead times for low precipitation probabilities (≤ 31). In summer and autumn (with the exception of the highest probability) EXIM outperforms persistence for all probabilities at all lead times. In winter the performance of EXIM is poorest out of all the seasons, where for higher probabilities (≥ 71) persistence outperforms EXIM. The improved performance of EXIM compared with PC can be attributed to the use of the IR10.8 channel and CT product in PCPh – in both products EXIM outperformed persistence.

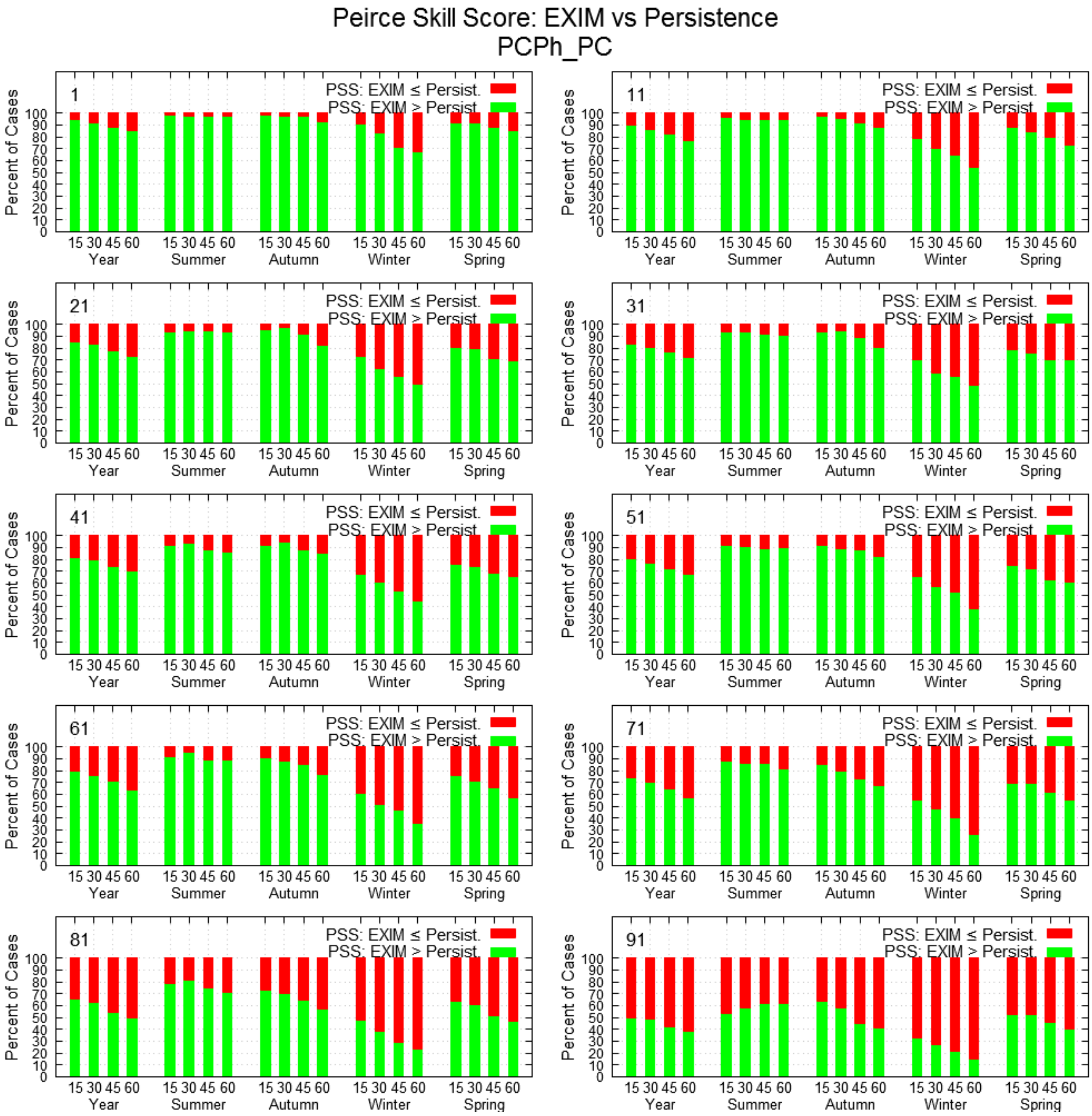

Figure 38: *As in Figure 3 but for Precipitating Clouds from Cloud Physical Properties Product. The threshold values correspond to precipitation likelihood with higher values corresponding to higher precipitation probabilities.*

2.13 SPhR: SEVIRI Physical Retrieval Products – Layer Precipitable Water

The SEVIRI Physical Retrieval (SPhR) products[2] use the SEVIRI WV6.2, WV7.3, IR8.7, IR10.8, IR12.0, and IR13.4 channels as input. As SPhR is a clear air product, it also requires as input the CMa to identify the clear air pixels.

The results of the forecasts of the Boundary Layer (BL), Middle Layer (ML), High Layer (HL), and the Total Precipitable Water (TPW) are presented in Figure 39, Figure 40, Figure 41, and Figure 42, respectively. The threshold values (*T*) represent values in the range [8,127] according to the definition of the SPhR product, and can be scaled to physical values (p) according to the equation:

$$p = \frac{T-8}{127-8} * (S_{max} - S_{min}) + S_{min}$$

where S_{max} and S_{min} are the maximum and minimum values that the product can have. These are defined in Table 6.

Parameter	Minimum	Maximum
SPhR_BL	0	35
SPhR_ML	0	45
SPhR_HL	0	8
SPhR_TPW	0	70

Table 6: *Minimum and maximum values for each of the moisture products.*

In general, EXIM performs poorer than persistence for BL, ML, and TPW, and better than persistence at later lead times for HL. As higher moisture content thresholds (≥ 70) occur infrequently, the number of cases (Figure 63 – Figure 66) need to be considered when interpreting the results. Unlike for the other products, the performance of EXIM, in general, improves with increasing lead time.

The poor performance of EXIM with SPhR_ML and SPhR_HL seems to contradict the good performance of the water vapour products (SEVIRI WV6.2 and WV7.3 µm) presented in Figure 17 and Figure 18, respectively.

A possible explanation for the observed results is a strong influence of the temporally interpolated NWP data in the calculation of the SPhR products. As a consequence, the evaluations favour those approaches closer resembling the interpolation method, and the actual behaviour of the atmosphere and its prediction are reflected in the statistics to a far lesser degree.

[2] SPhR was the designation in v2013 (for which the validation was carried out). The product has been renamed in version NWC/GEO v2016 as iSHAI (Imaging Satellite Humidity and Instability).

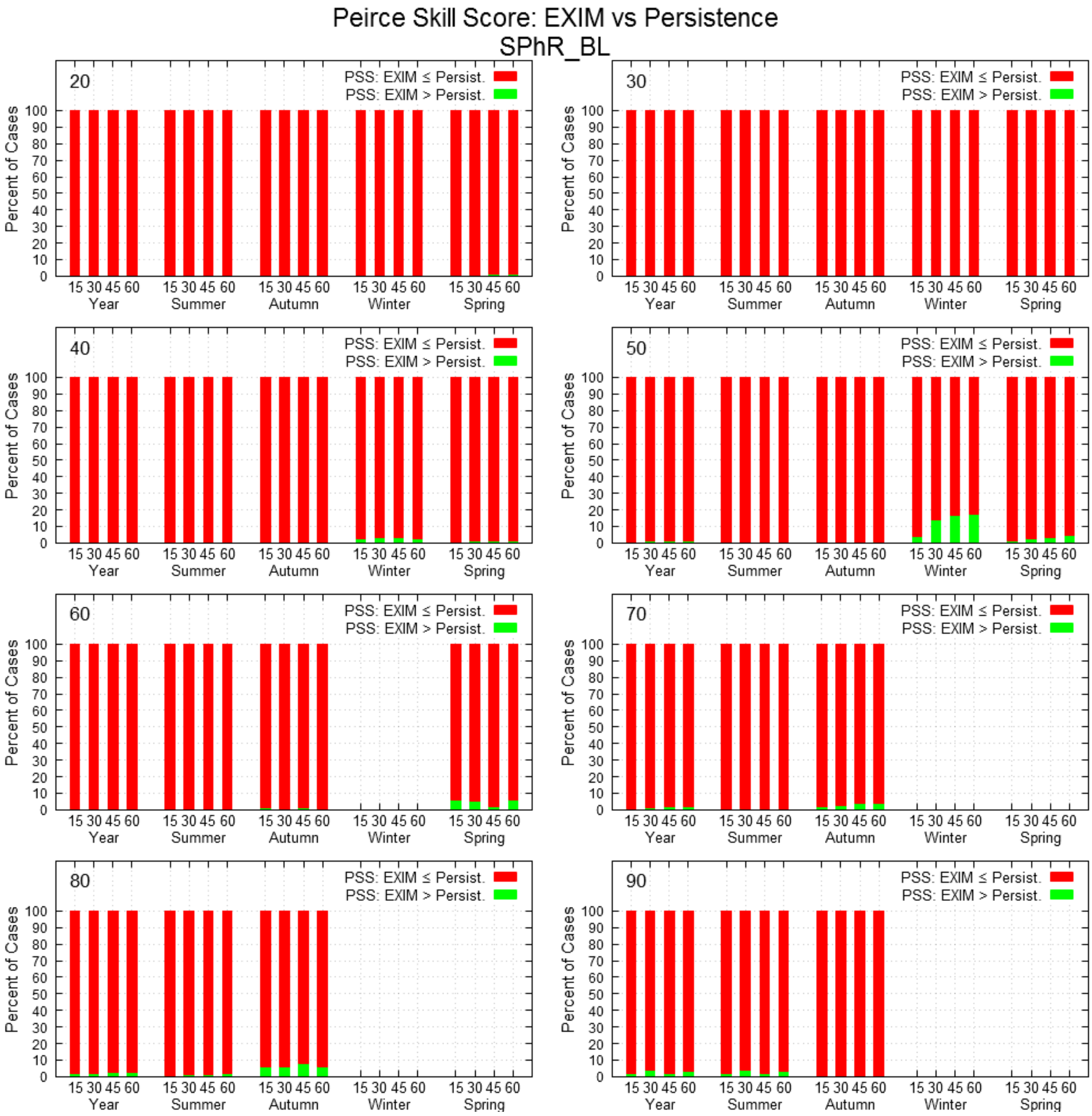

Figure 39: *As in Figure 3 but for SPhR_BL Product.*

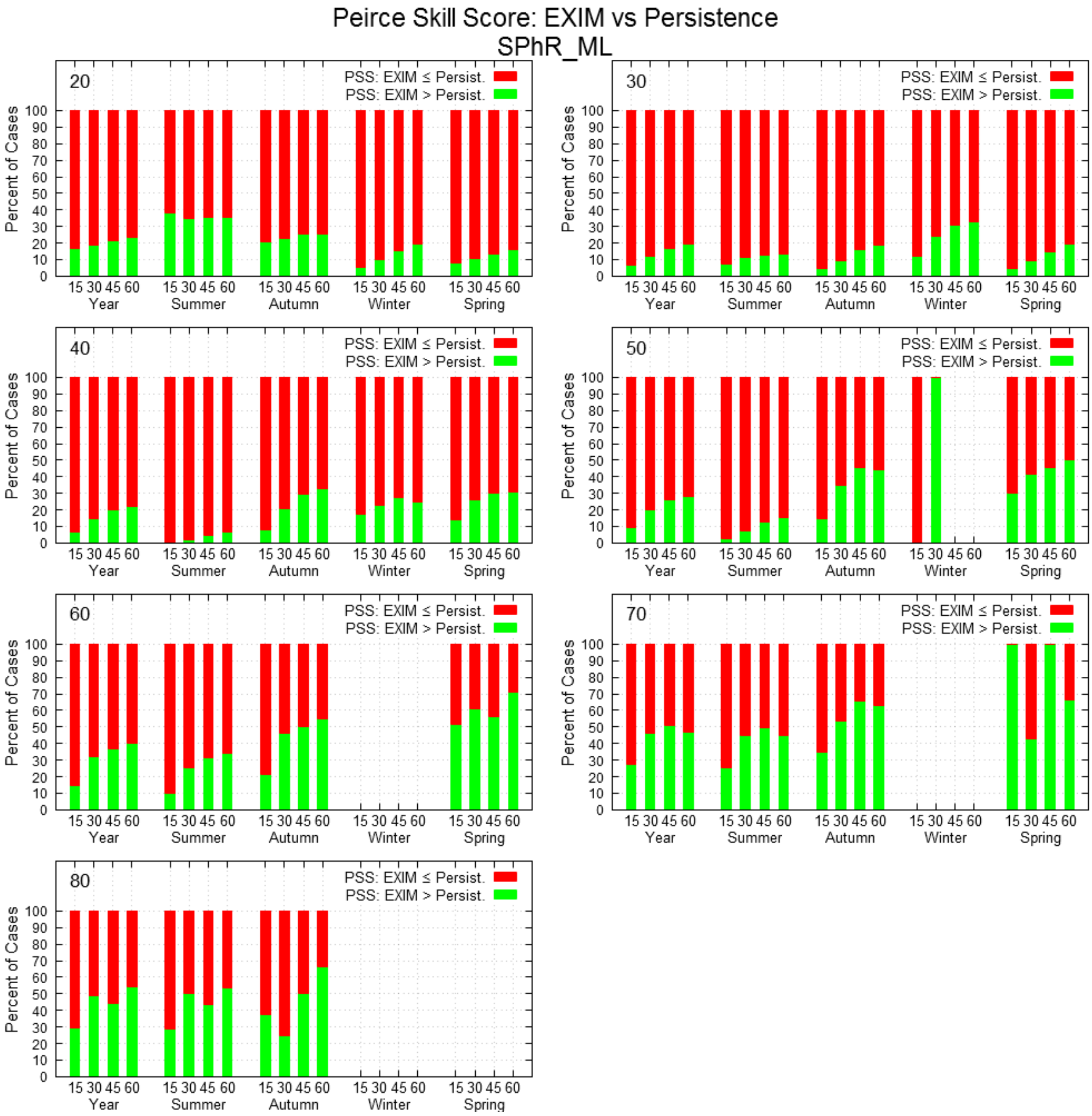

Figure 40: *As in Figure 3 but for SPhR_ML Product.*

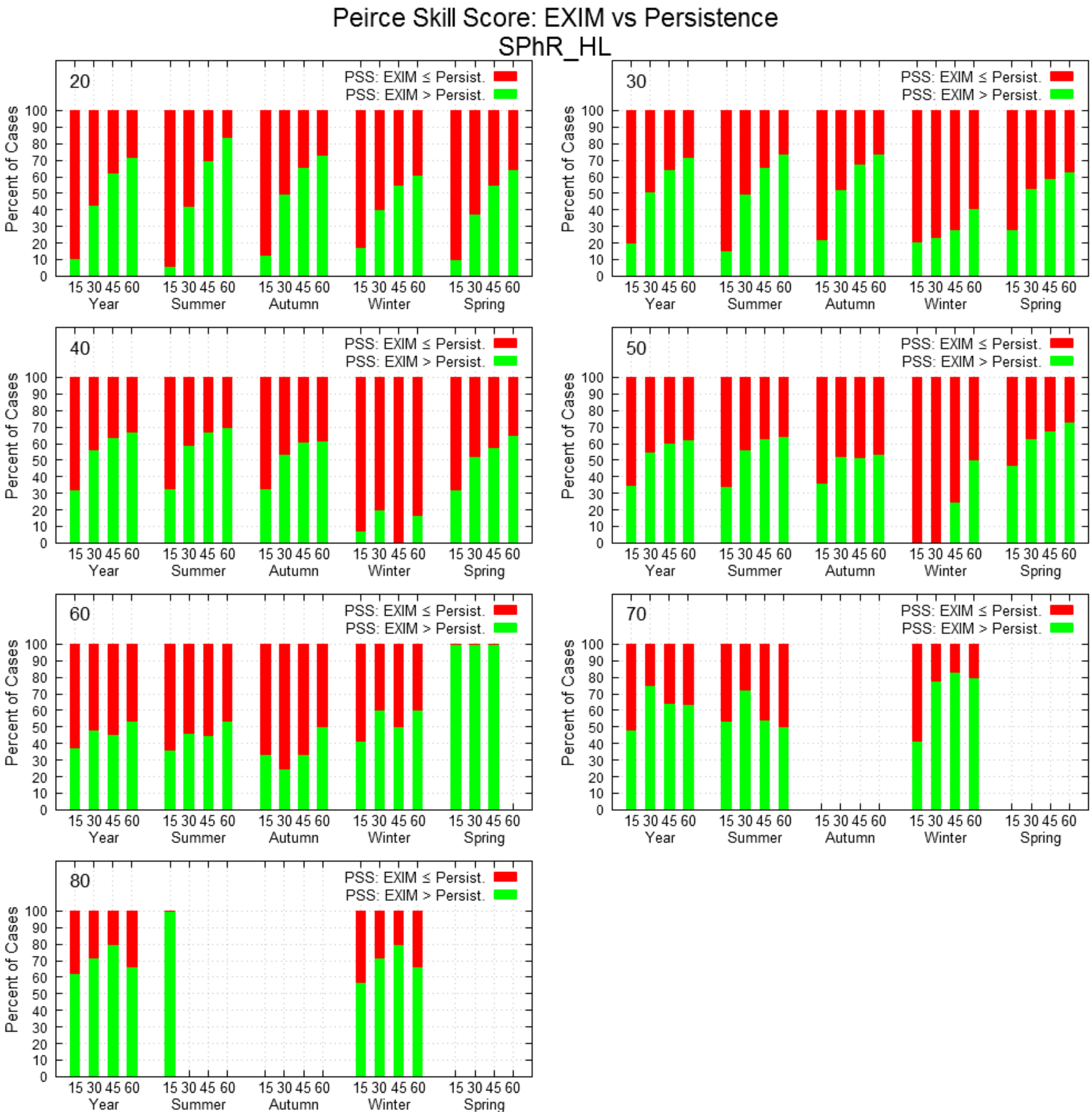

Figure 41: *As in Figure 3 but for SPhR_HL Product.*

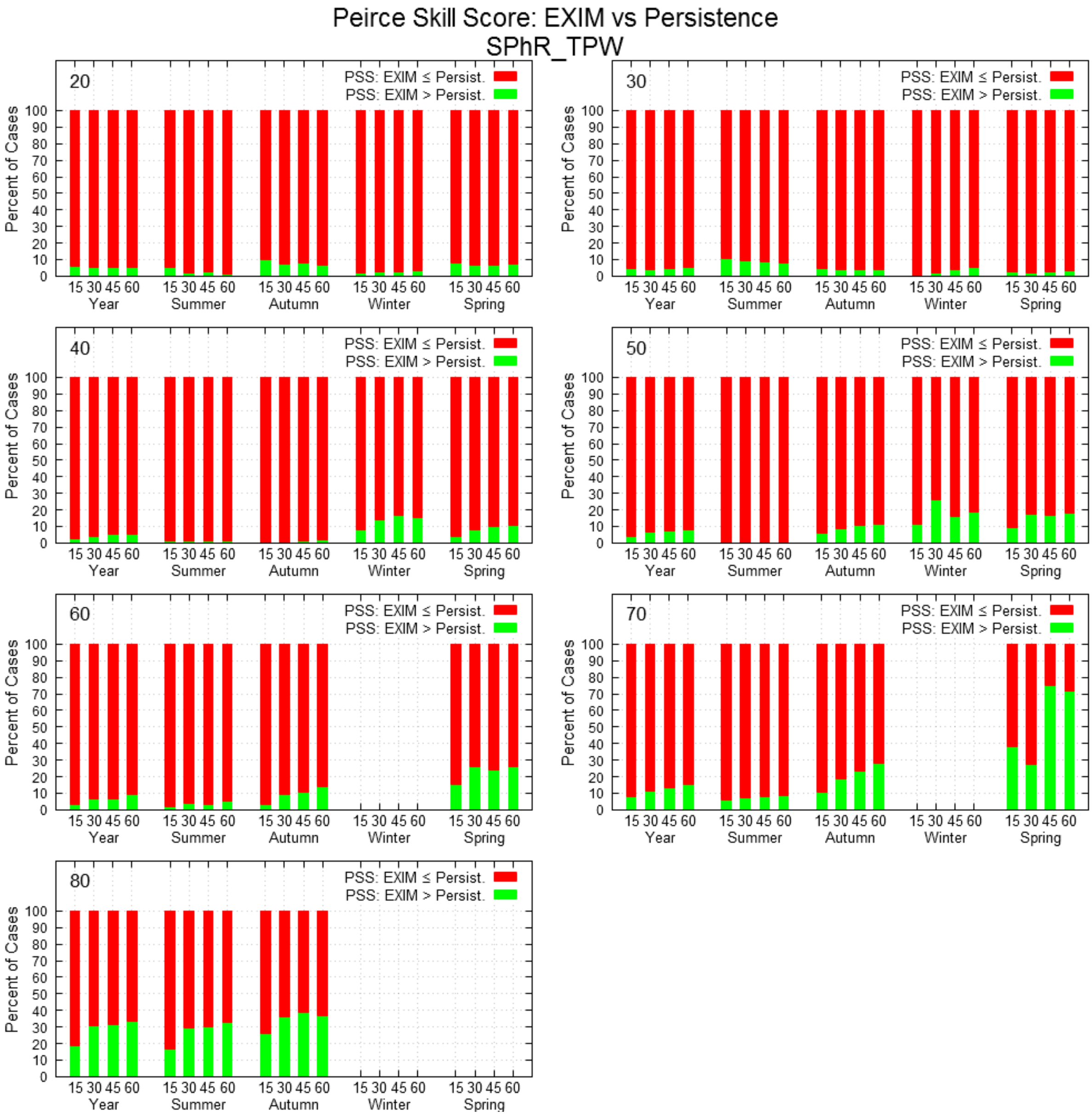

Figure 42: As in Figure 3 but for TPW Product.

2.14 SPhR: SEVIRI Physical Retrieval Products – Stability Analysis

The results of the forecasts of the K-Index (KI), Lifted Index (LI), and Showalter Index (SHW) are presented in Figure 43, Figure 44, and Figure 45, respectively. As described for the SPhR moisture products, the threshold values (*T*) represent values in the range [8,127] and can be converted to physical values by the equation given for *p* previous. The corresponding values of S_{max} and S_{min} are defined in Table 7. Larger threshold values represent larger atmospheric stabilities.

In general, EXIM performs poorer than persistence for all stability products regardless of threshold, season, and lead time. The same suspicion on the overwhelming influence of the temporally interpolated NWP data as in section 2.13 holds.

Parameter	**Minimum**	**Maximum**
SPhR_KI	0	60
SPhR_LI	-15	25
SPhR_SHW	-15	25

Table 7: *Minimum and maximum values for each of the stability products.*

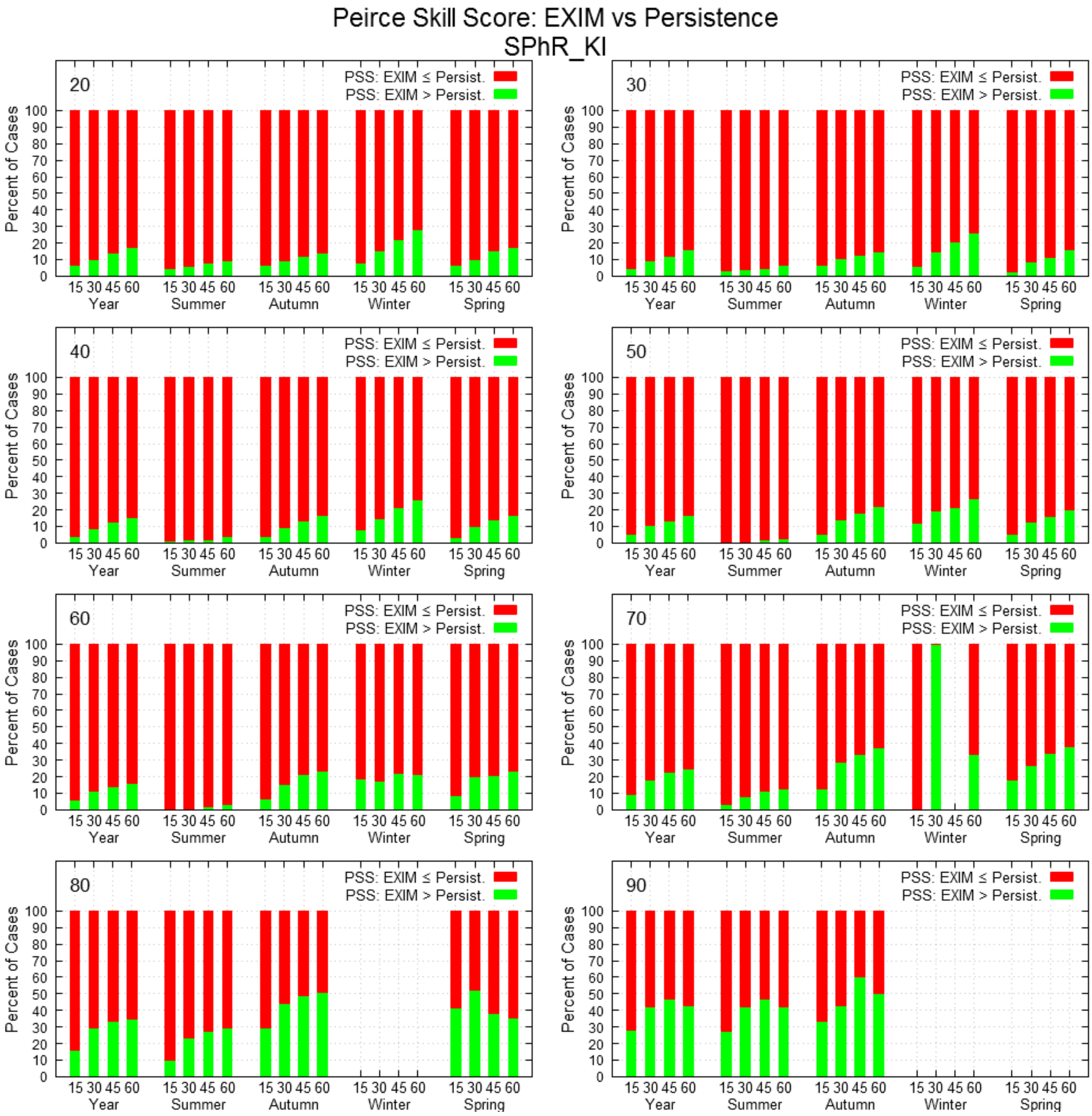

Figure 43: *As in Figure 3 but for SPhR_KI Product.*

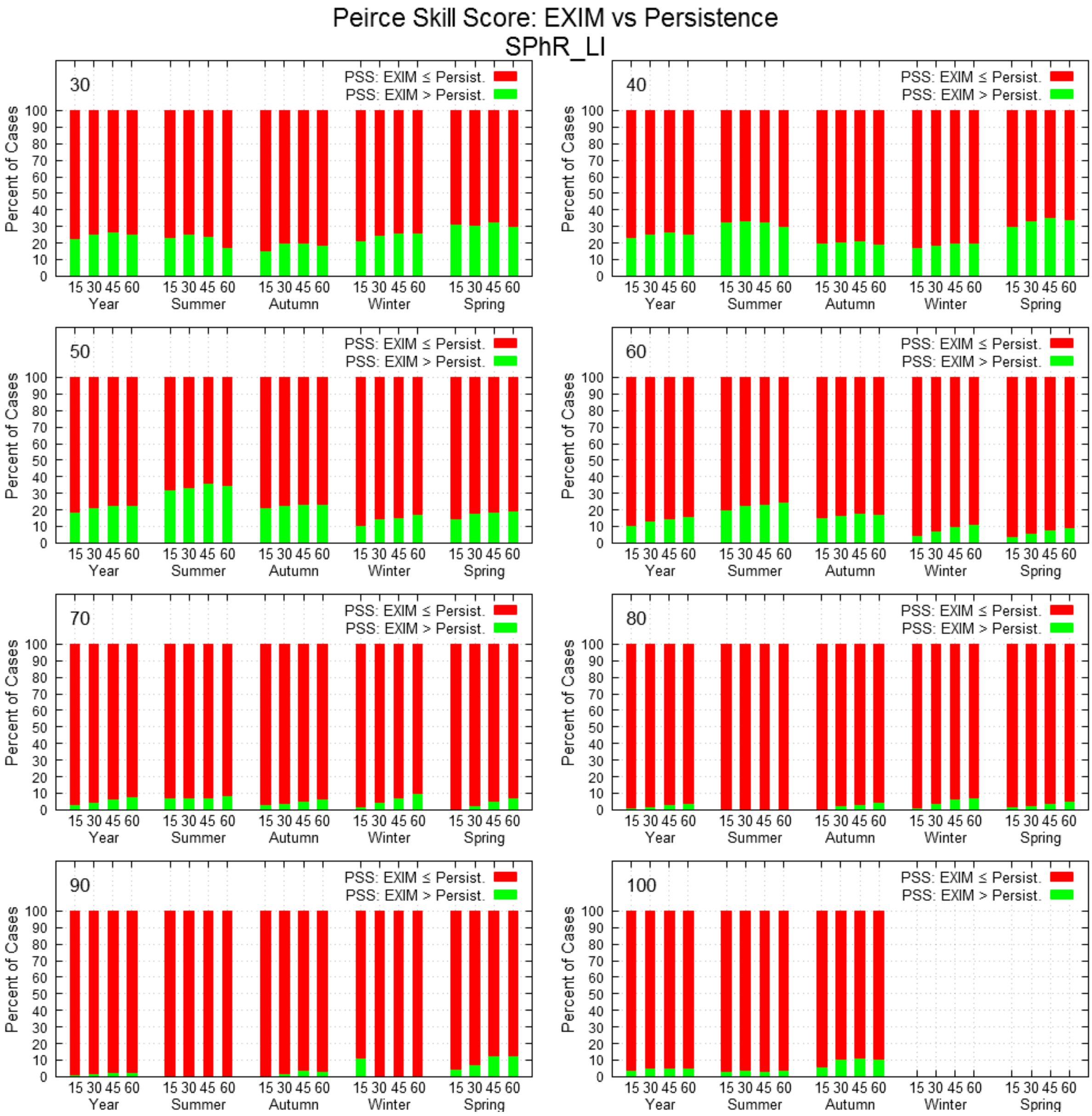

Figure 44: As in Figure 3 but for SPhR_LI Product.

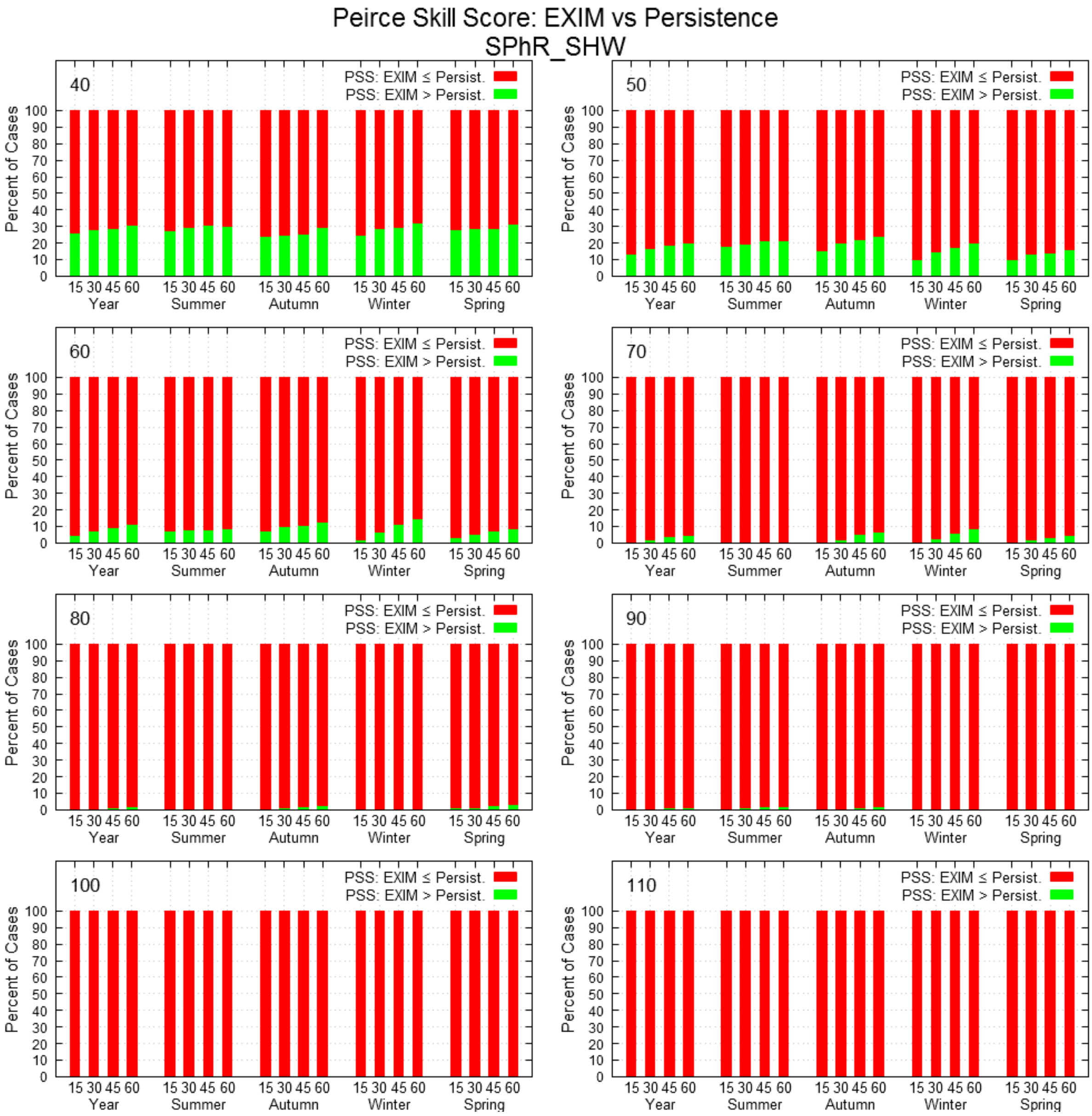

Figure 45: *As in Figure 3 but for SPhR_SHW Product.*

3. TABULAR SUMMARY OF VALIDATION RESULTS

The following section analyses the PSS results of each product shown in the previous section to provide recommendations where EXIM can be used to provide skilful forecasts. In this context, skilful forecasts mean that EXIM outperforms persistence, i.e. PSS(EXIM) > PSS(Persistence). In general, EXIM is recommended when the number of cases for which it outperforms persistence is greater than 50% for all seasons and all lead times, unless otherwise stated.

3.1 Thermal Channels: Infrared

Product	Use of EXIM	Exceptions
IR3.9	Summer: BT ∈ [250, 280] K Autumn, Spring: BT ∈ [250, 270] K Winter: BT ∈ [250, 270] K	- - 270 K, LT +60 min
IR8.7	Summer: BT ∈ [230, 280] K Autumn, Spring: BT ∈ [230, 270] K Winter: BT ∈ [230, 270] K	- - 270 K, LT +60 min
IR9.7	Summer: BT ∈ [230, 250] K Autumn, Spring: BT ∈ [230, 250] K Winter: BT ∈ [230, 240] K	- - -
IR10.8	Summer, Autumn: BT ∈ [230, 280] K Spring: BT ∈ [230, 280] K Winter: BT ∈ [230, 270] K	- - 280 K, LT ≥ +45 min
IR12.0	Summer, Autumn: BT ∈ [230, 280] K Spring: BT ∈ [230, 280] K Winter: BT ∈ [230, 270] K	- - 280 K, LT ≥ +45 min
IR13.4	Summer, Autumn: BT ∈ [230, 260] K Spring: BT ∈ [230, 260] K Winter: BT ∈ [230, 250] K	- - 260 K, LT +60 min

3.2 Thermal Channels: Water Vapour

Product	Use of EXIM	Exceptions
WV6.2	BT ∈ [220, 240] K	-
WV7.3	BT ∈ [220, 260] K	-

3.3 Visible Channels

Product	Use of EXIM	Exceptions
VIS0.6	Radiances ∈ [4, 8]	Spring, Autumn: LT +60 min
VIS0.8	Radiances = 6 - Radiances = 8	Summer, Autumn: LT +60 min, Spring: LT ≥ +45 min Spring: LT +60 min
NIR1.6	Not recommended	

It should be noted that, since data for the SEVIRI visible channels were not available for winter, a recommendation for the usage of EXIM in winter could not be made.

3.4 CT: Cloud Type

Product	Use of EXIM	Exceptions
CT	Recommended for all seasons and lead times	Winter: LT +60 min

3.5 CTTH: Cloud Top Temperature and Height

Product	Use of EXIM
CTTH	Recommended for CTTH ≤ 8000 m

3.6 CRR: Convective Rainfall rate

Product	Use of EXIM
CRR	Recommended for CRR ≤ 5 mm/h

3.7 CRPh: Convective Rainfall Rate from Cloud Physical Properties

Product	Use of EXIM	Exceptions
CRPh	CRPh ≤ 11 (2.2 mm/h)	-
	Summer: CRPh ≤ 71 (14.2 mm/h)	-
	Autumn: CRPh ≤ 71 (14.2 mm/h)	61, 71: LT +60 min
	Winter: CRPh ≤ 41 (8.2 mm/h)	LT +60 min
	Winter: CRPh ≤ 71 (14.2 mm/h)	LT ≥ 45 min
	Spring: CRPh ≤ 41 (8.2 mm/h)	LT +60 min
	Spring: CRPh ≤ 61 (12.2 mm/h)	LT ≥ 45 min

3.8 PC: Precipitating Clouds

Product	Use of EXIM	Exceptions
PC	Summer: ≤ 60 %	+15 min at 40 % and 60 %
	Autumn ≤ 30 %	-
	Winter = 20%	-
	Spring ≤ 30 %	10%

3.9 PCPh: Precipitating Clouds from Cloud Physical Properties

Product	Use of EXIM	Exceptions
PCPh	Recommended for PCPh ≤ 31%	Winter, LT +60 min for PCPh ≥ 21 %

3.10 SPhR: SEVIRI Physical Retrieval Products – Layer Precipitable Water

Product	Use of EXIM
SPhR_BL	Not recommended
SPhR_ML	Not recommended
SPhR_HL	Not recommended
SPhR_TPW	Not recommended

3.11 SPhR: SEVIRI Physical Retrieval Products – Stability Analysis

Product	Use of EXIM
SPhR_KI	Not recommended
SPhR_LI	Not recommended
SPhR_SHW	Not recommended

4. CONCLUSIONS

EXIM utilises calculated atmospheric motion vectors (AMVs) to extrapolate the motion of features from satellite data, or products generated from satellite data, to produce nowcasts in 15 minute intervals out to one hour. The success of the forecasts from EXIM has been evaluated by comparing them with what was observed. As a basis to determine how good EXIM is, the persistence forecast is also compared with the observations. The success of both EXIM and Persistence are evaluated using the Peirce Skill Score (PSS), with higher values corresponding to a better performance. The dataset used covers a year from 17 September 2014 to 17 September 2015, although a dataset from 8 April 2015 to 17 September 2015 was available for the SEVIRI visible products. The dataset was split according to season and the performance of both forecast methods for each season assessed.

The forecasts of EXIM for the infrared channels (3.9 µm, 8.7 µm, 9.7 µm, 10.8 µm, 12.0 µm, and 13.4 µm) are better than those from Persistence, with the exception being for winter at the highest brightness temperature (BT) threshold at +60 minutes forecast lead time. During winter the highest brightness temperatures correspond with snow/ice on the ground, which should not be moved.

The forecasts of EXIM for the water vapour channels (6.2 µm, 7.3 µm) are better than those from Persistence at the BTs considered. However, the forecasts of EXIM for the visible and near-infrared channels (0.6 µm, 0.8 µm, 1.6 µm) are in general poorer than the Persistence forecasts, due to ground features being detected.

The forecasts of EXIM for Cloud Type are better than Persistence for all seasons and lead times, with the exception of winter at the lead time of +60 minutes.

The forecasts of EXIM for Cloud Top Temperature and Height (CTTH) are better than Persistence for all seasons but only for cloud tops at or below 8000 m. The reason for this may be due to the increasing blocky nature of the CTTH field at height levels above 8000 m. According to the CTTH documentation [RD.3], it is possible to configure the product such that it exploits pixel resolution throughout (at the cost of increased computation time). For the next evaluation campaign, this will have to be implemented in order to obtain a more complete and fairer picture.

The forecasts of EXIM for both Convective Rainfall Rate (CRR) and Convective Rainfall Rate from Cloud Physical Properites (CRPh) are better than Persistence at lower precipitation rates for all seasons. For moderate precipitation rates the forecasts of EXIM are generally better than Persistence for lead times +15 min and +30 min.

The forecasts of EXIM for Precipitating Clouds (PC) and Precipitating Clouds from Cloud Physical Properties (PCPh) are better than Persistence for low probabilities depending on the season.

The forecasts of EXIM for the SEVIRI Physical Retrieval Products involving both Layer Precipitable Water and Stability Analysis are worse than those from Persistence. As a consequence, these products were dropped from the EXIM portfolio.

The performance of EXIM can be improved through a number of measures. One such measure includes preprocessing the input data using a cloud mask to eliminate displacement of surface features. This improvement should lead to better results during winter for the infrared channels with high BTs, and for all visible and near-infrared channels. At the time of writing, a test implementation already exists at ZAMG and is expected to be released in NWC/GEO v2018 after successful validation (or as a patch even beforehand). Another obvious candidate for future exploration is the use of height-varying displacement fields, with two provisos: 1) the computation time constraints on the NWC/GEO package have to allow for that, also in the MTG resolution; 2) successful strategies have to be developed for those situations where the high-level and low-level displacement fields markedly differ, which would lead to large gaps in the resulting image or, conversely, systems from different layers converging to one place in the image.

As mentioned earlier, the SAFNWC Product Requirements Table in [AD.8] defines the threshold accuracy of EXIM as "*on average better than persistence forecast*". The target accuracy is formulated as "*always better than persistence forecast*". This is the minimum requirement to release the product; the tables in section 3 showed the term "not recommended" where the evaluation indicated performance below this level. The target accuracy is to be "*always better than persistence forecast*"; the optimal accuracy (which by construction is almost impossible to achieve) is described as "all advective changes are perfectly captured". Non-advective changes are one argument to justify a certain tolerance about the term "always" when judging whether target accuracy is reached or not: in cases of newly developing or strongly decaying systems, neither EXIM nor persistence may capture the near future very well, and the sign of PSS(EXIM)-PSS(Persistence) is at random. Hence, for verifying on target accuracy, we rather check here whether EXIM performs better in the vast majority of cases, and thus arrive at the following judgements:

Parameter	***Threshold accuracy reached?***	***Target accuracy reached?***
SEVIRI IR Channels	Yes	Yes over most of the temperature range; for the warmest temperatures, incorporation of cloud mask to suppress extrapolation in cloud free areas seems necessary
SEVIRI WV channels	Yes	Yes over most of the temperature range
SEVIRI VIS channels	Yes	No. Incorporation of cloud mask to suppress extrapolation in cloud free areas seems necessary
SEVIRI NIR1.6	No	No
CT	Yes	Yes for leadtime 15 minutes
CTTH	Yes	Yes for heights below 8000m; need to resolve "blocky nature" issue for fair evaluation in higher levels
PC	Yes	No
PC-Ph	Yes	No (with some hopes associated with improved illumination handling in forthcoming PC-Ph release)
CRR	Yes	No
CRR-Ph	Yes	No (with some hopes associated with improved illumination handling in forthcoming CRR-Ph release)
SPhR (renamed in version NWC/GEO v2016 as iSHAI (Imaging Satellite Humidity and Instability)	No	No

A. ANNEX: COMPARISON PLOTS IN TERMS OF NUMBER OF CASES

Plots showing the performance of EXIM compared with persistence in terms of number of cases observed.

A.1 THERMAL CHANNELS: INFRARED

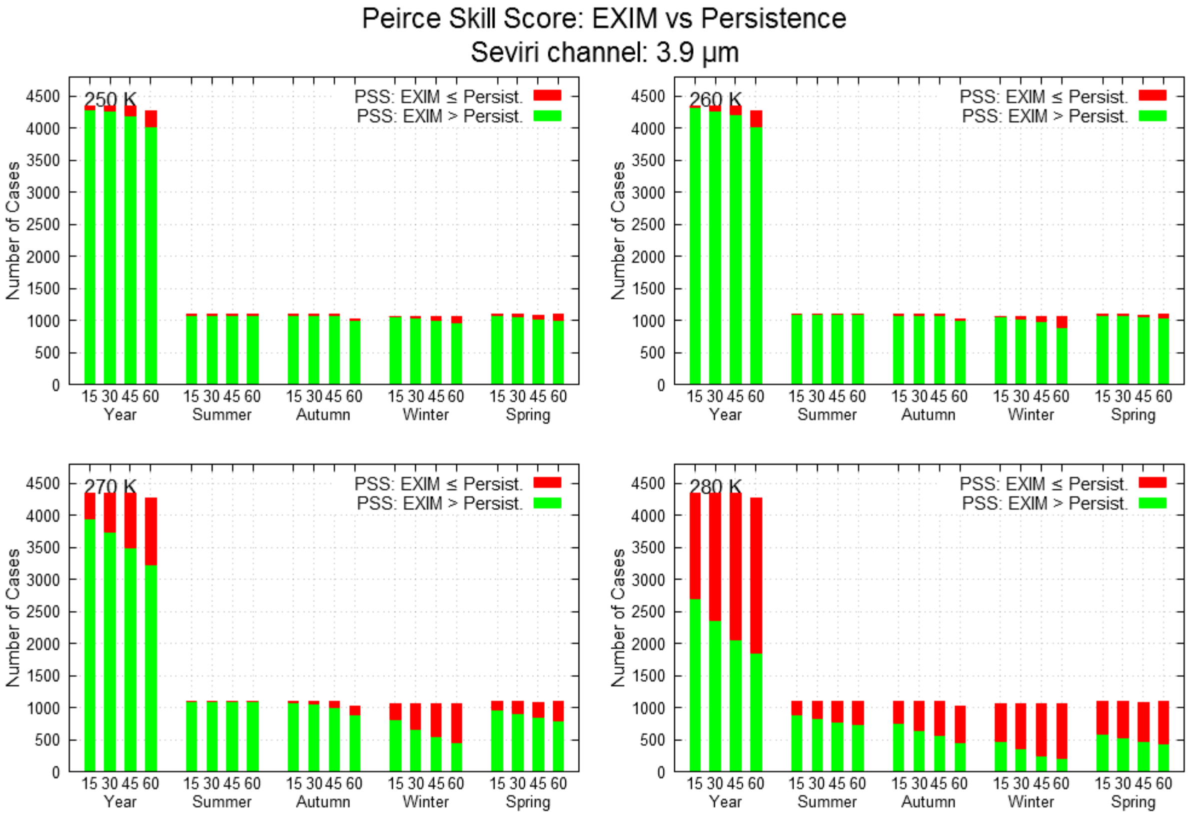

Figure 46: *SEVIRI channel 3.9µm. Histogram showing the number of occurrences where the Peirce Skill Score (PSS) of EXIM exceeds (green), or is less than or equal to (red) that for persistence. Each panel corresponds to a threshold value shown in the upper left. Each group of bins represent the data according to season along with a sum over the year, and is separated into the lead times (+15, +30, +45, +60 minutes).*

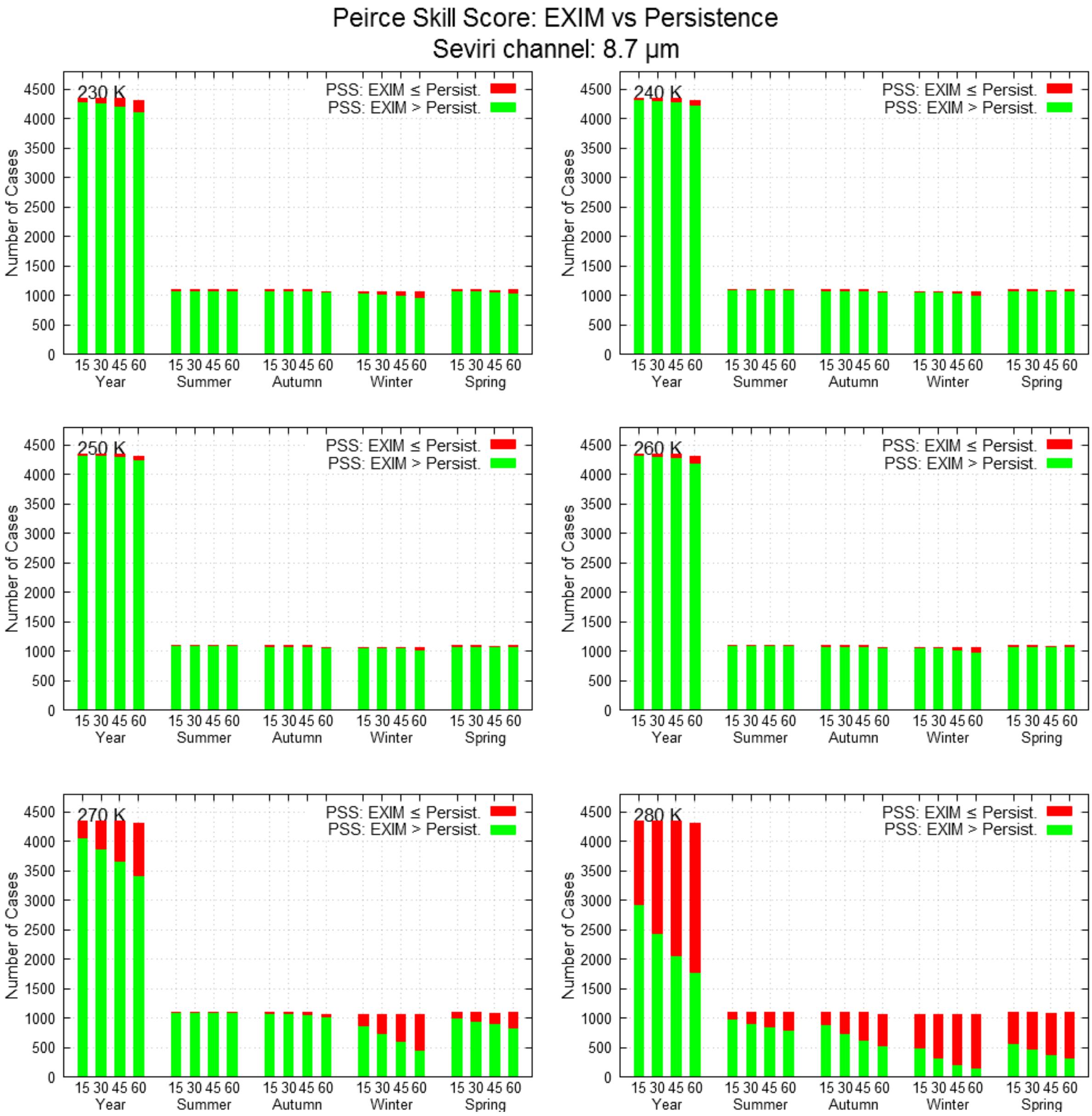

Figure 47: *As in Figure 46 but for SEVIRI channel 8.7µm.*

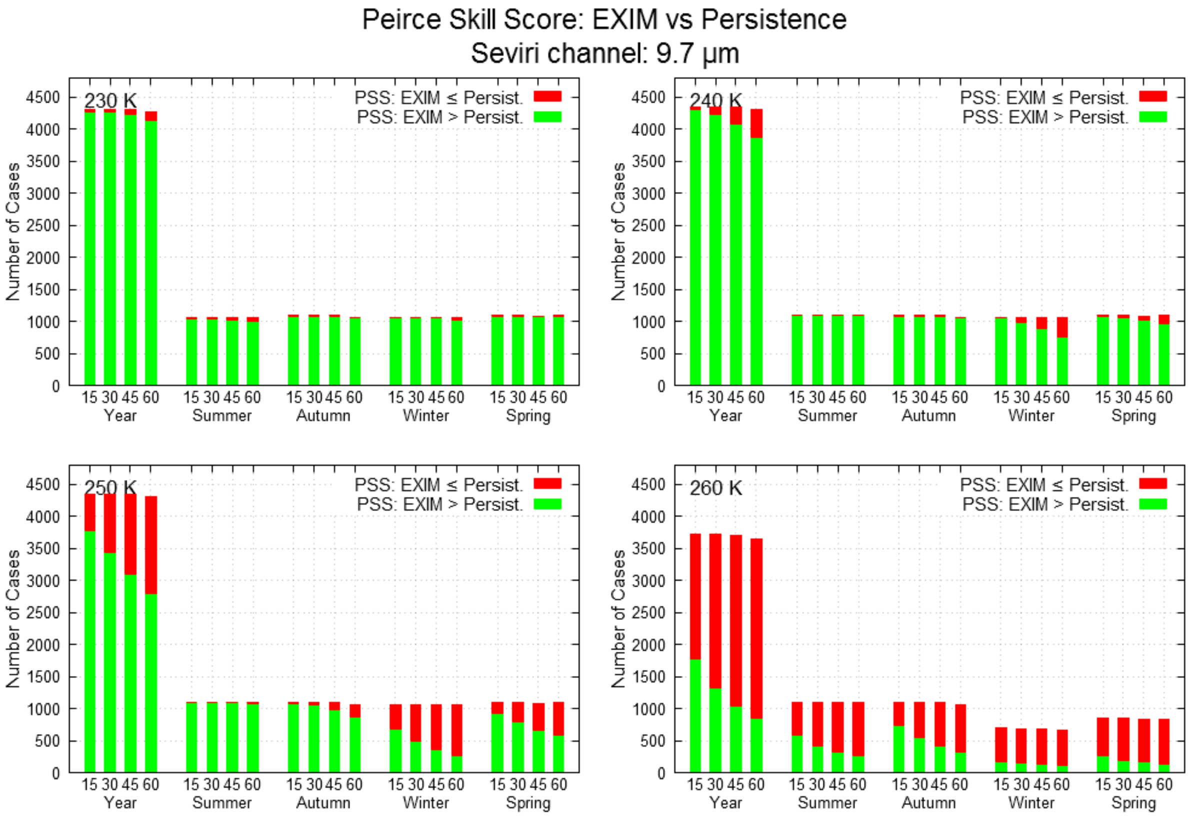

Figure 48: *As in Figure 46 but for SEVIRI channel 9.7µm.*

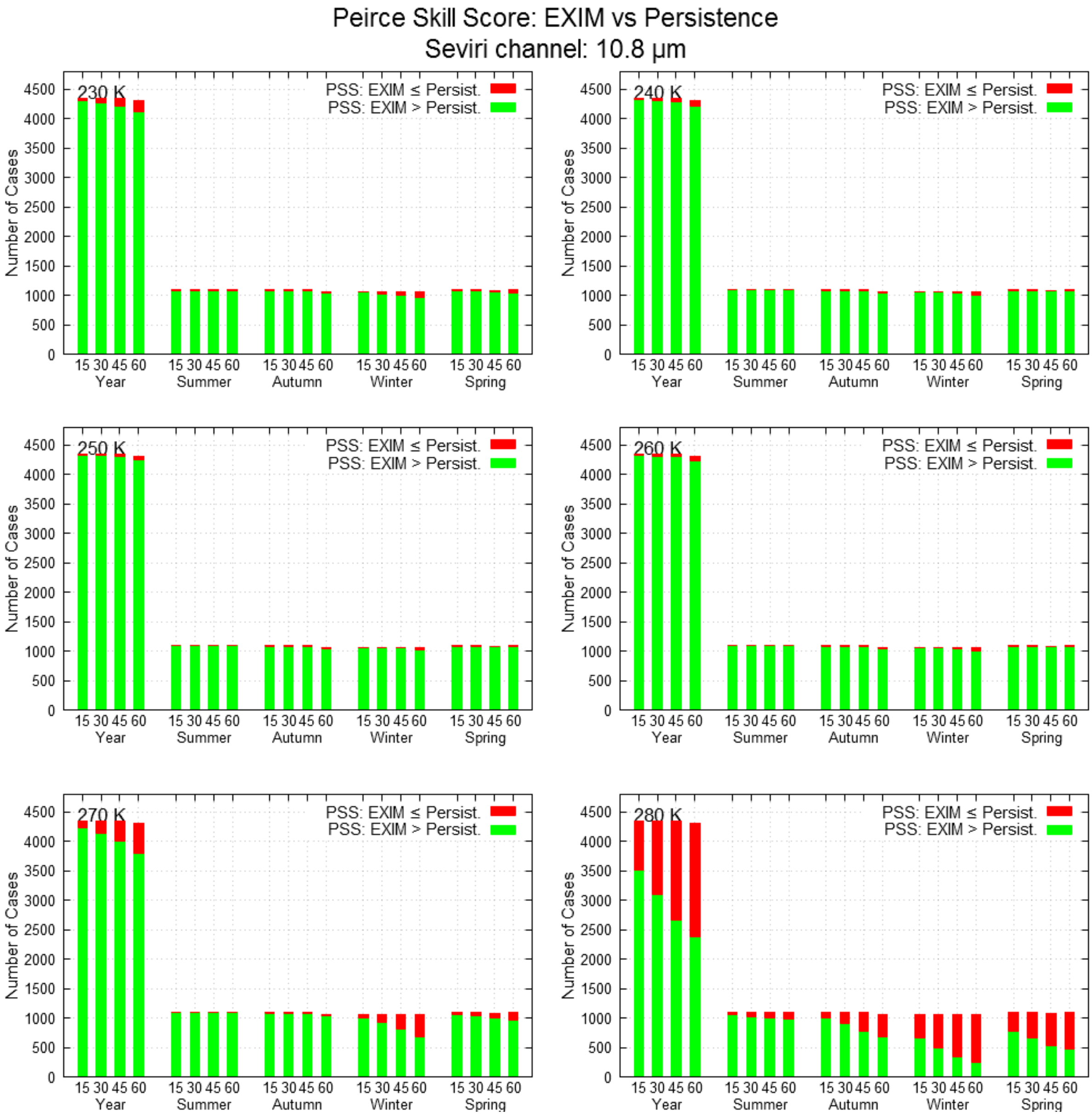

Figure 49: *As in Figure 46 but for SEVIRI channel 10.8µm.*

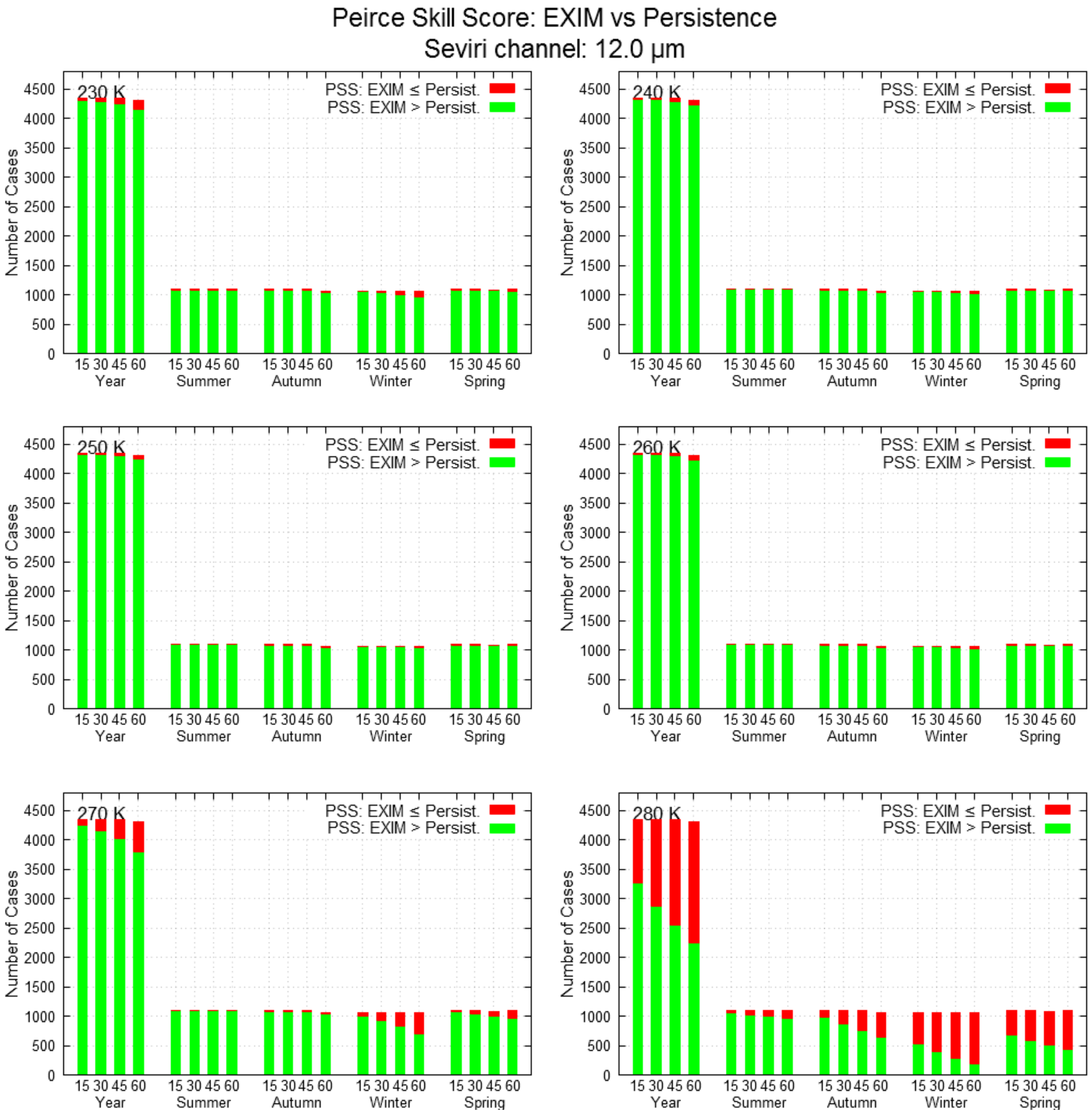

***Figure 50:** As in Figure 46 but for SEVIRI channel 12.0µm.*

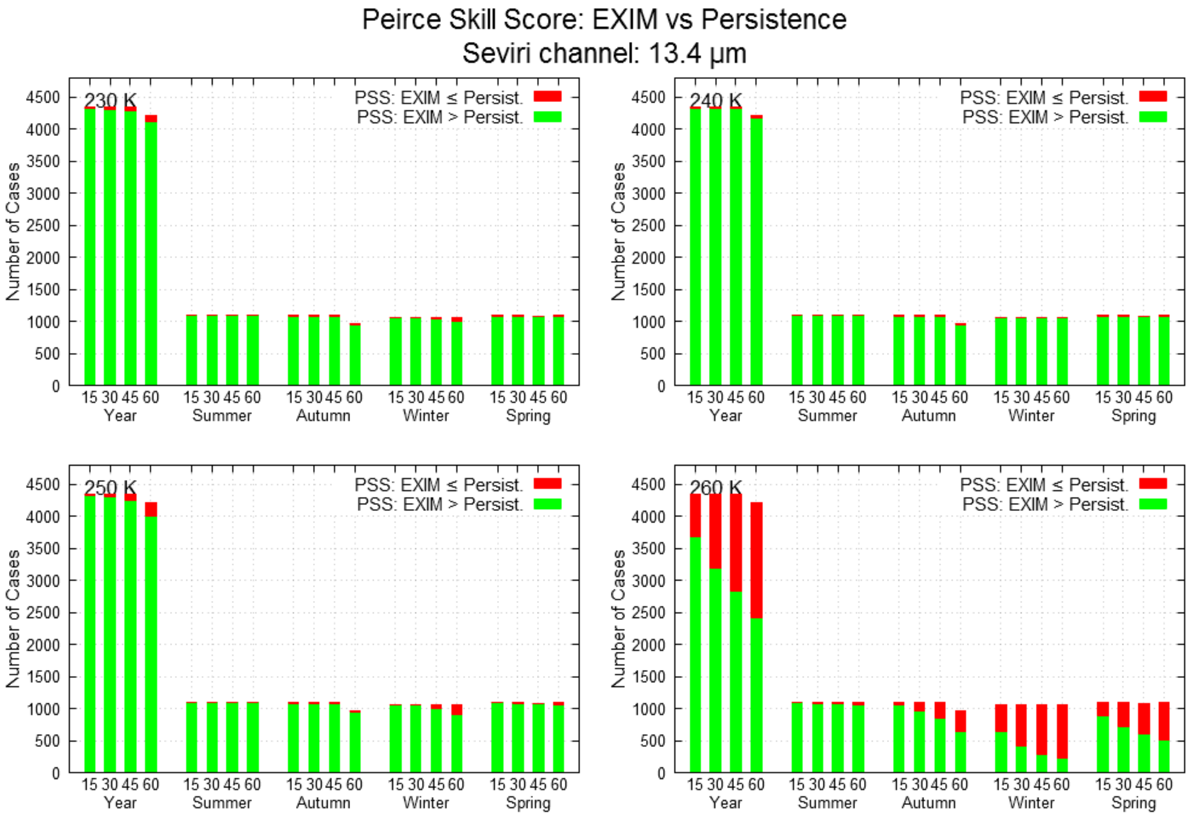

Figure 51: *As in Figure 46 but for SEVIRI channel 13.4µm.*

A.2 THERMAL CHANNELS: WATER VAPOUR

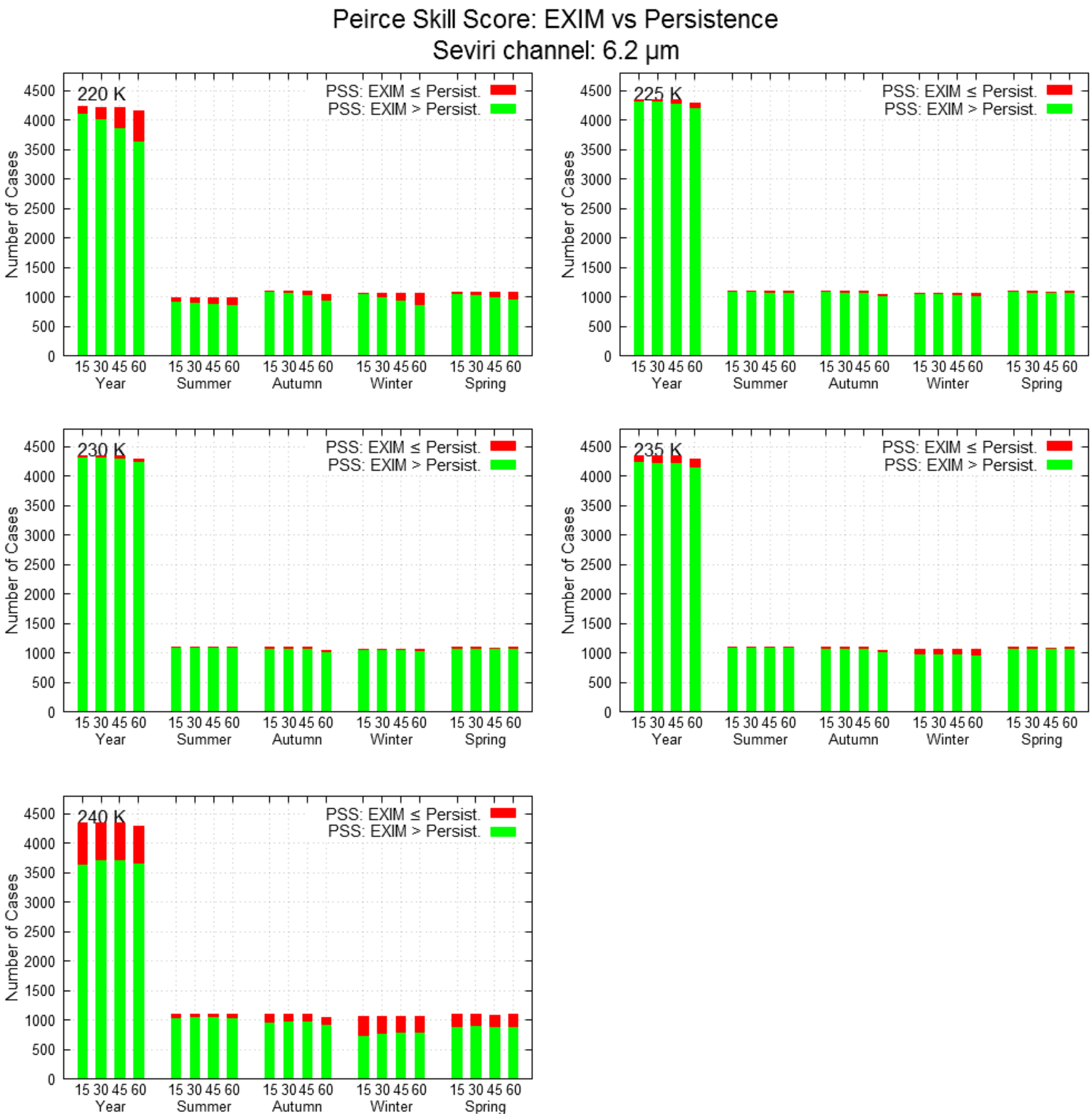

Figure 52: *As in Figure 46 but for SEVIRI channel 6.2µm.*

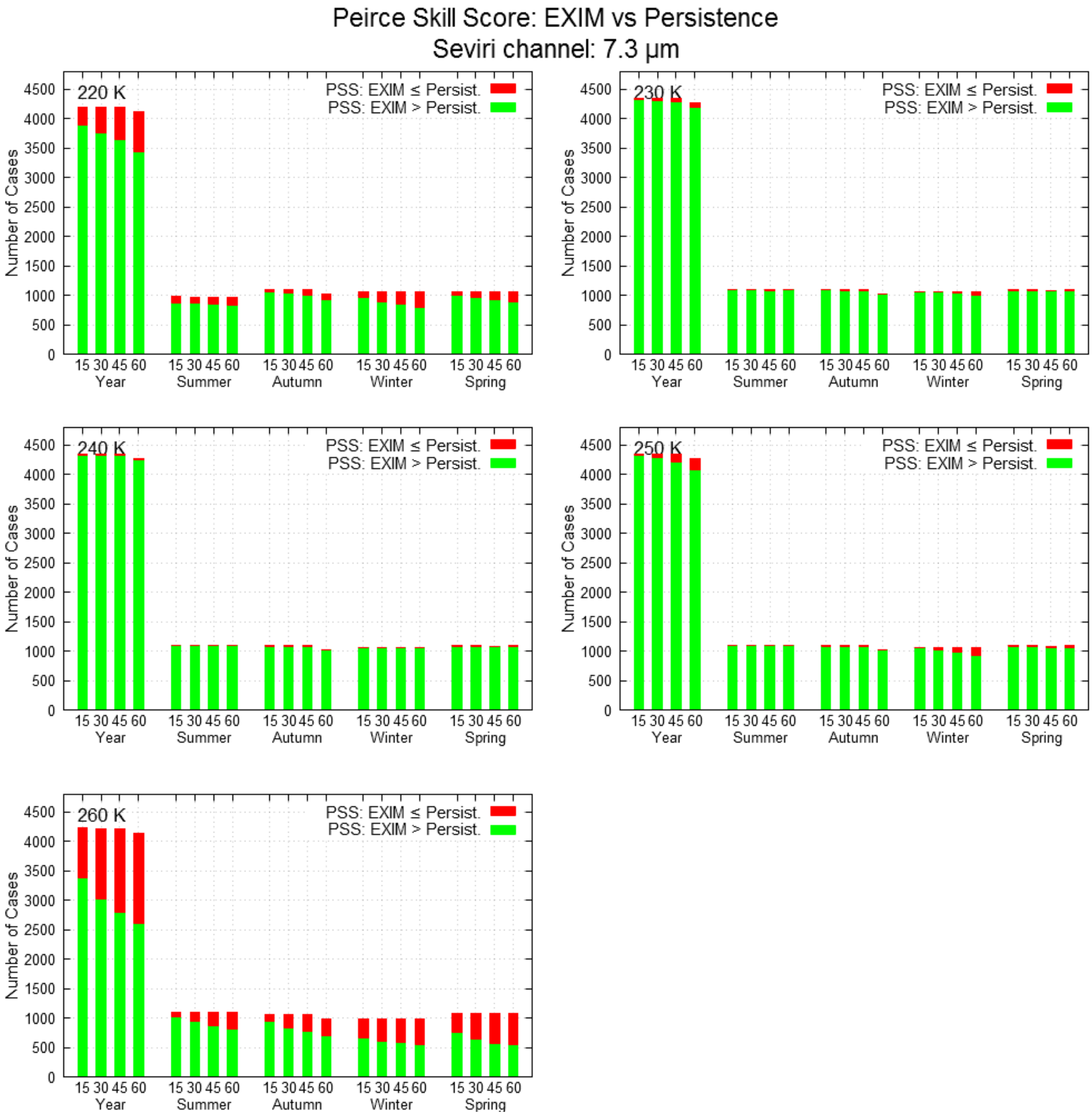

Figure 53: *As in Figure 46 but for SEVIRI channel 7.3µm.*

A.3 VISIBLE CHANNELS

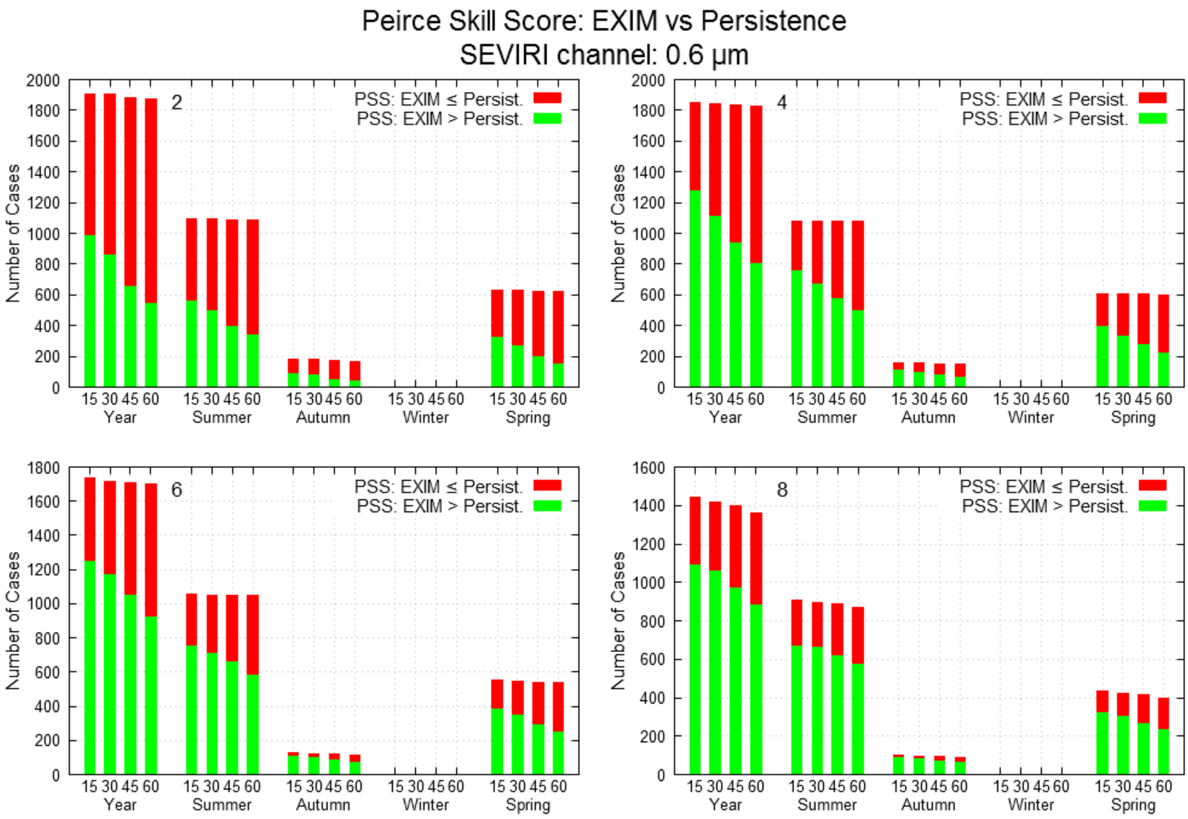

Figure 54: As in Figure 46 but for SEVIRI channel 0.6µm.

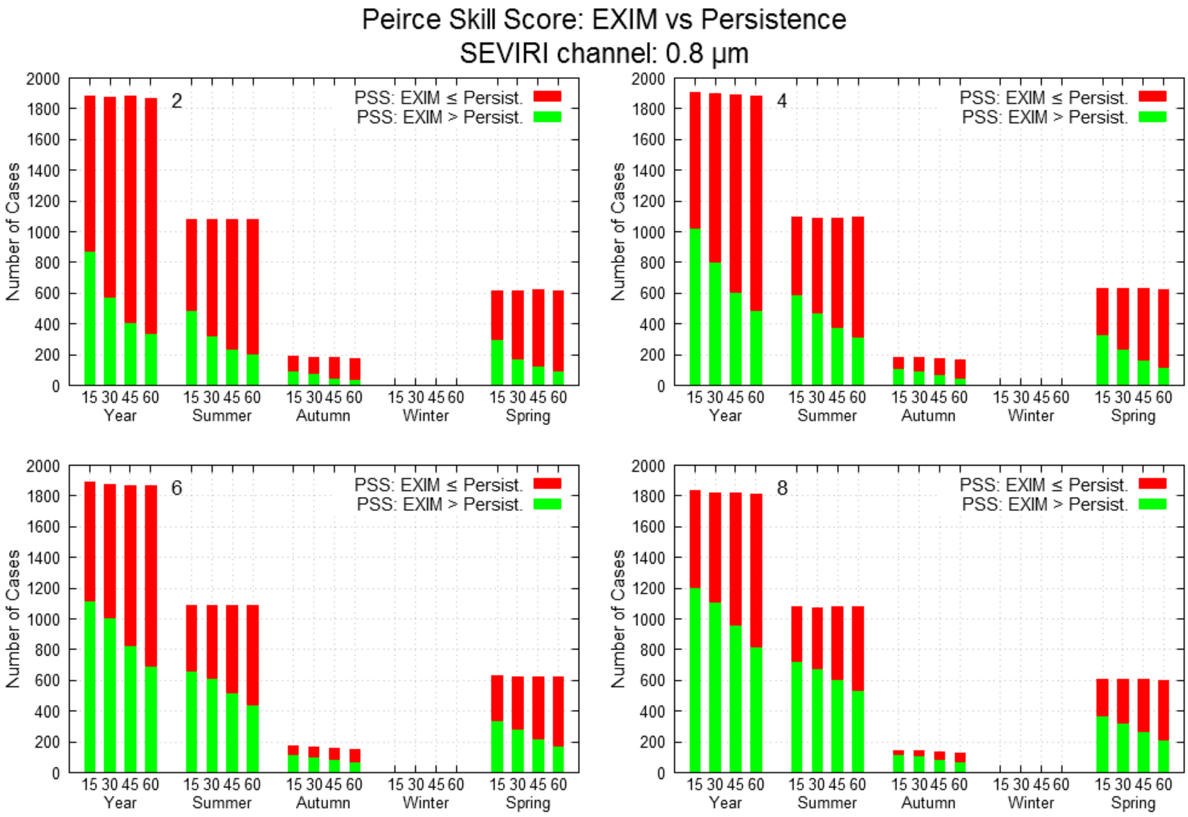

Figure 55: *As in Figure 46 but for SEVIRI channel 0.8µm.*

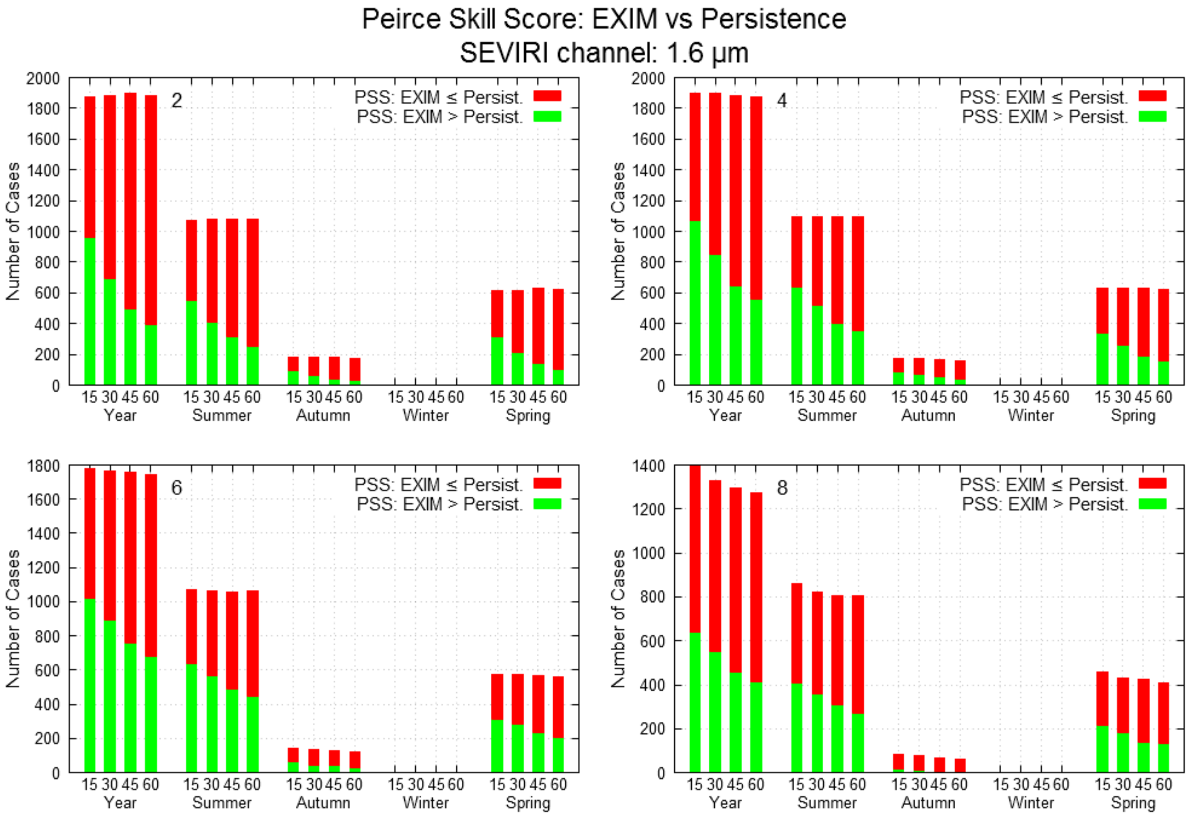

Figure 56: As in Figure 46 but for SEVIRI channel 1.6µm.

A.4 CT: CLOUD TYPE

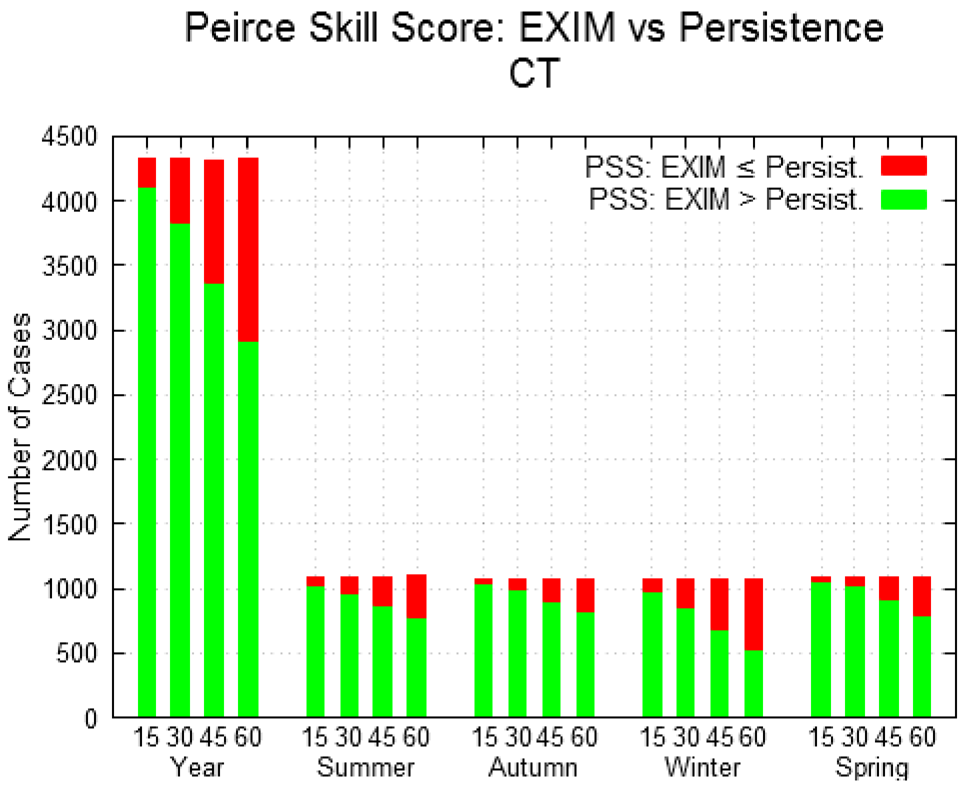

Figure 57: *As in Figure 46 but for Cloud Type.*

A.5 CTTH: CLOUD TOP TEMPERATURE AND HEIGHT

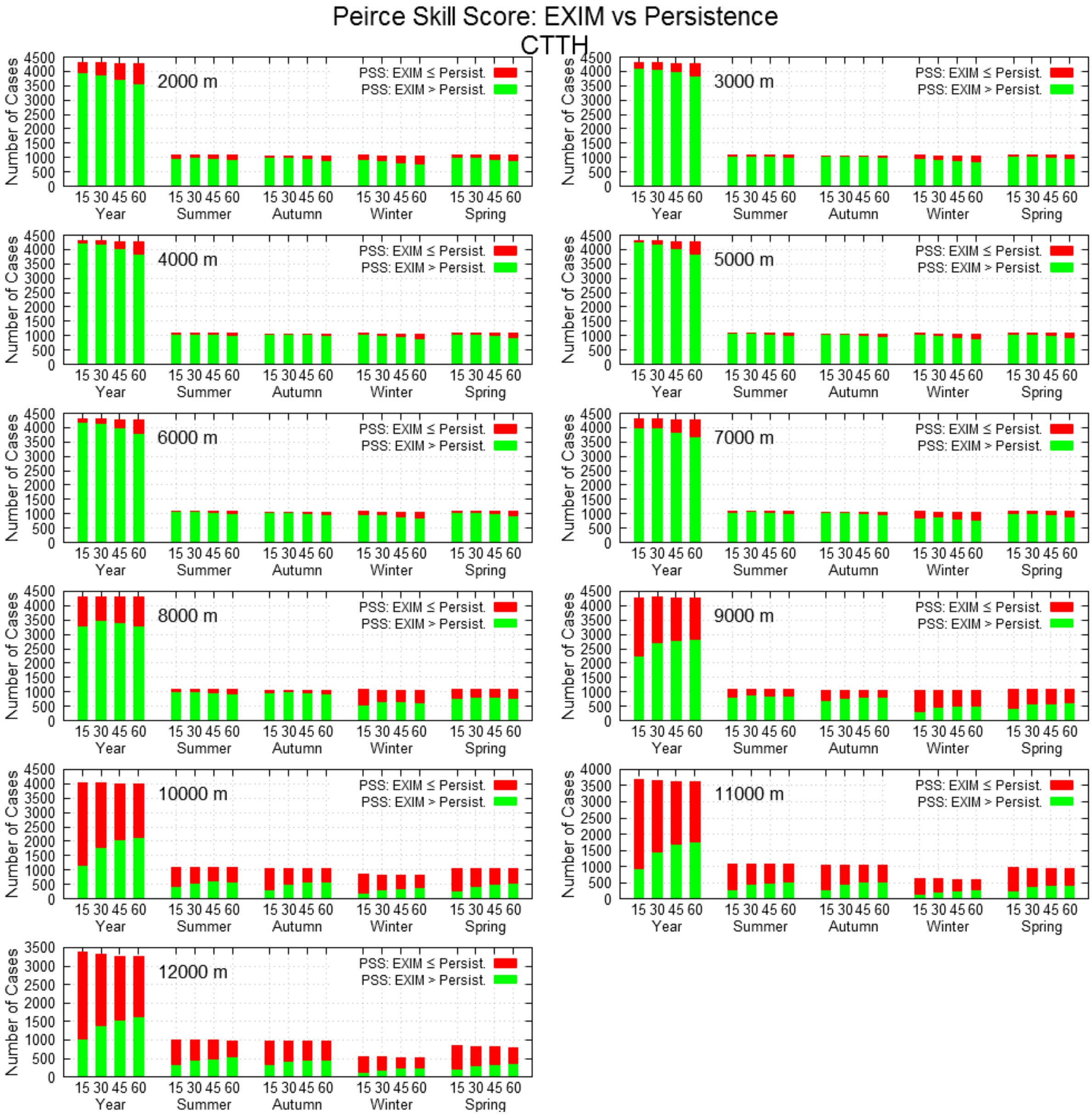

***Figure 58:** As in Figure 46 but for Cloud Top Temperature and Height Product.*

A.6 CRR: CONVECTIVE RAINFALL RATE

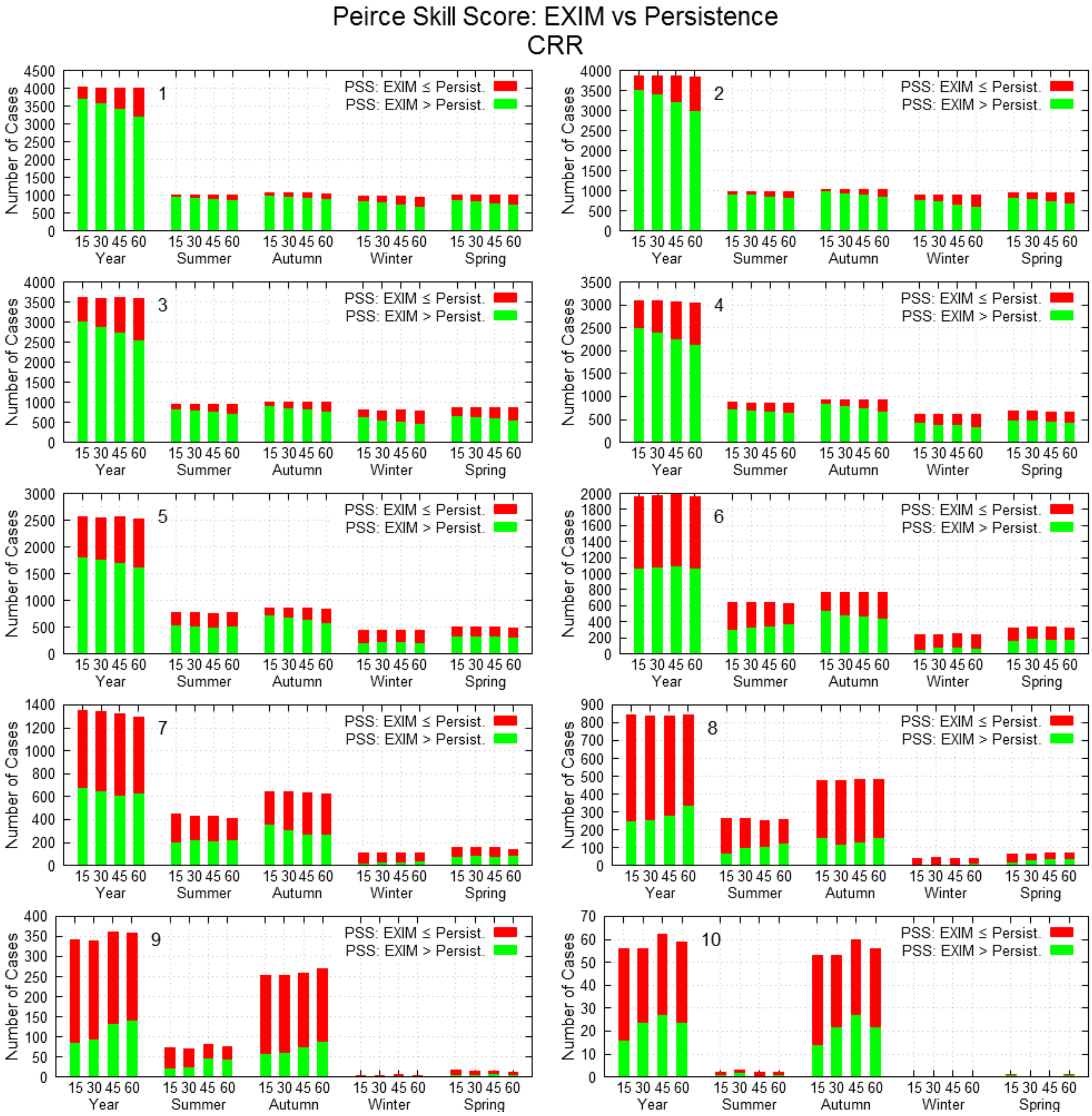

Figure 59: *As in Figure 46 but for Convective Rainfall Rate Product.*

A.7 CRPh: Convective Rainfall Rate from Cloud Physical Properties

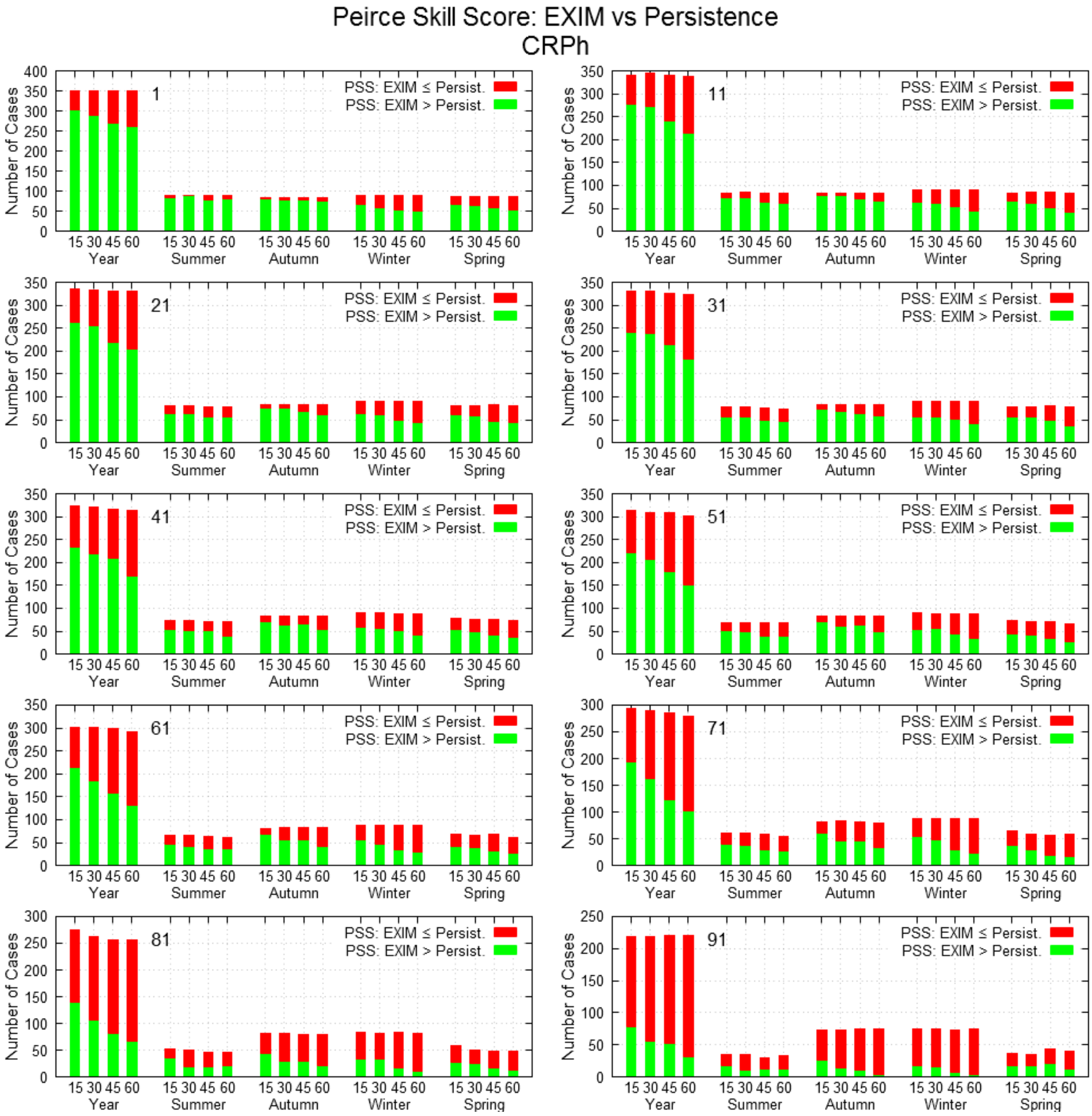

Figure 60: *As in Figure 46 but for Convective Rainfall Rate from Cloud Physical Parameters Product.*

A.8 PC: Precipitating Clouds

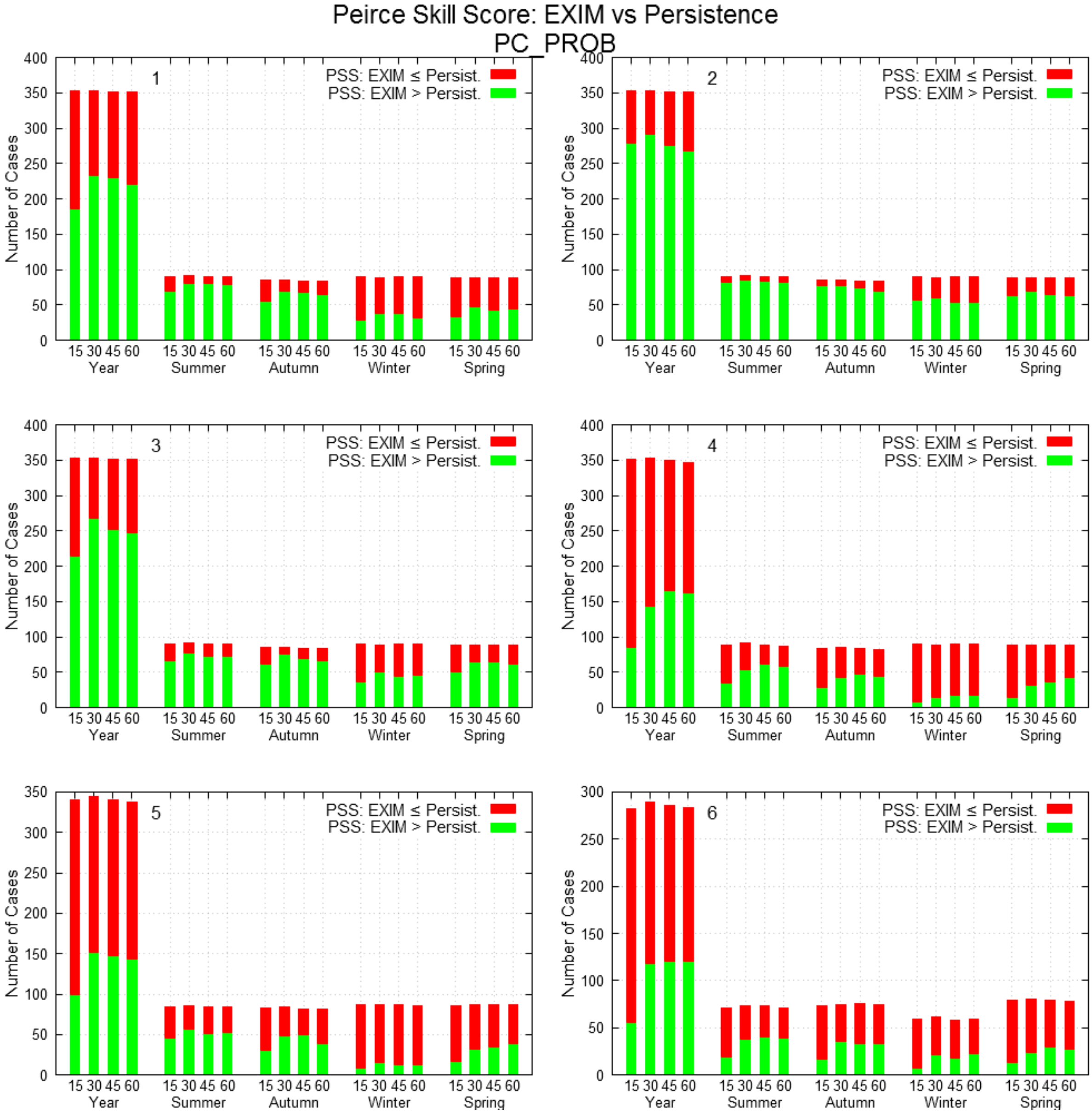

Figure 61: As in Figure 46 but for Precipitating Clouds Product.

A.9 PCPH: PRECIPITATING CLOUDS FROM CLOUD PHYSICAL PROPERTIES

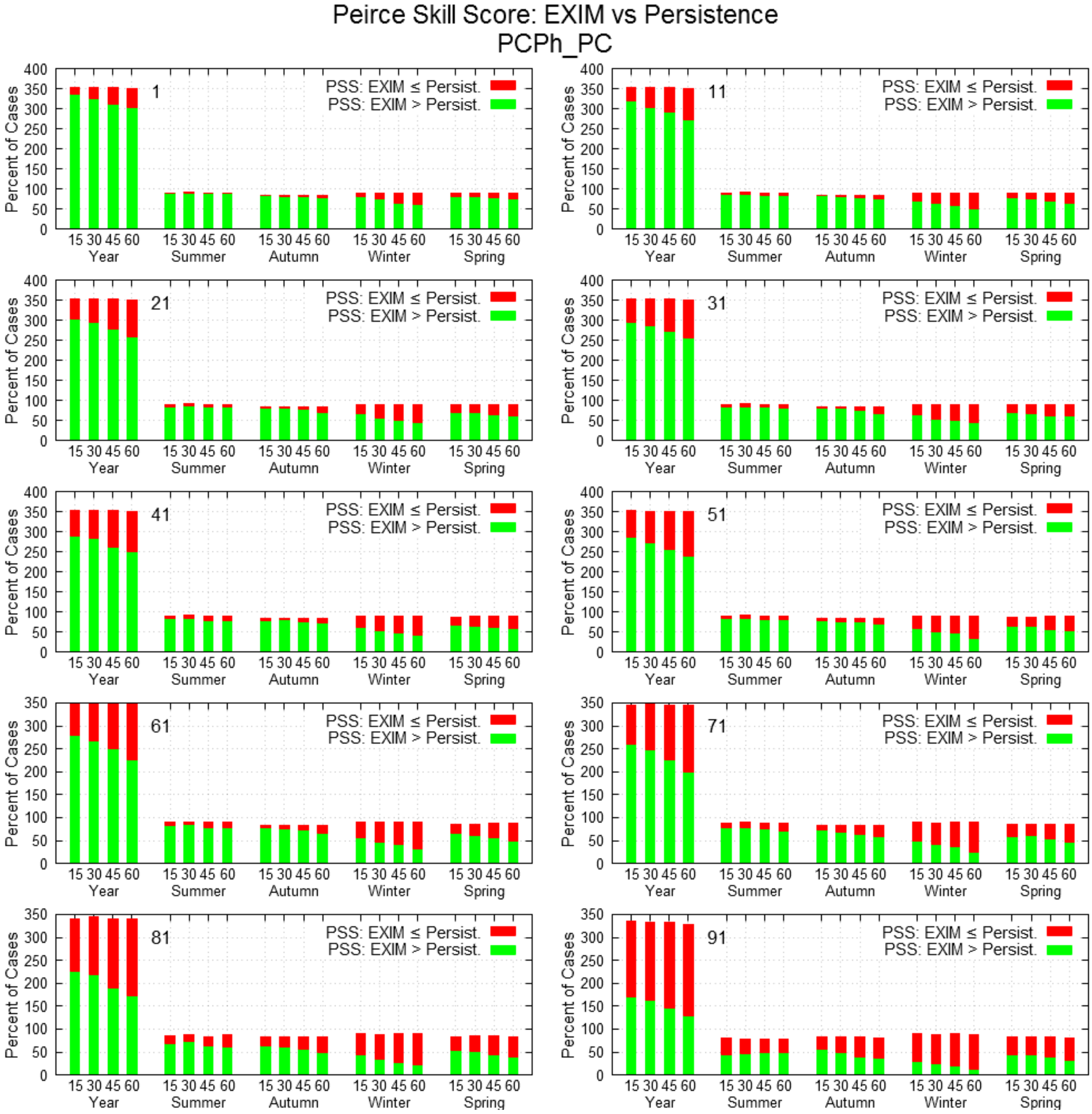

Figure 62: *As in Figure 46 but for Precipitating Clouds from Cloud Physical Properties Product.*

A.10 SPhR: SEVIRI Physical Retrieval Products[3] – Layer Precipitable Water

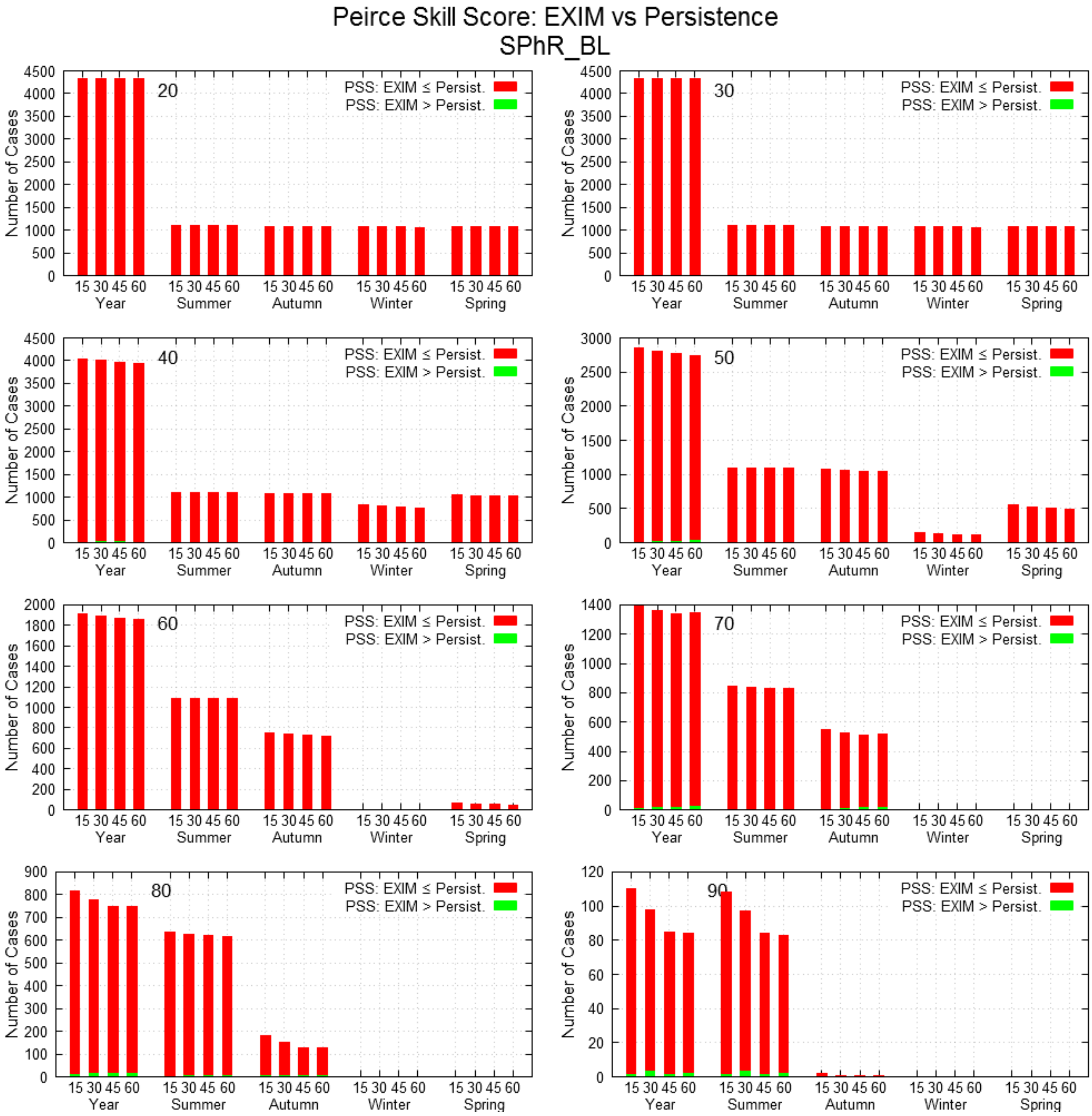

Figure 63: As in Figure 46 but for SPhR_BL Product.

[3] SPhR was the designation in v2013 (for which the validation was carried out). The product has been renamed in version NWC/GEO v2016 as iSHAI (Imaging Satellite Humidity and Instability)

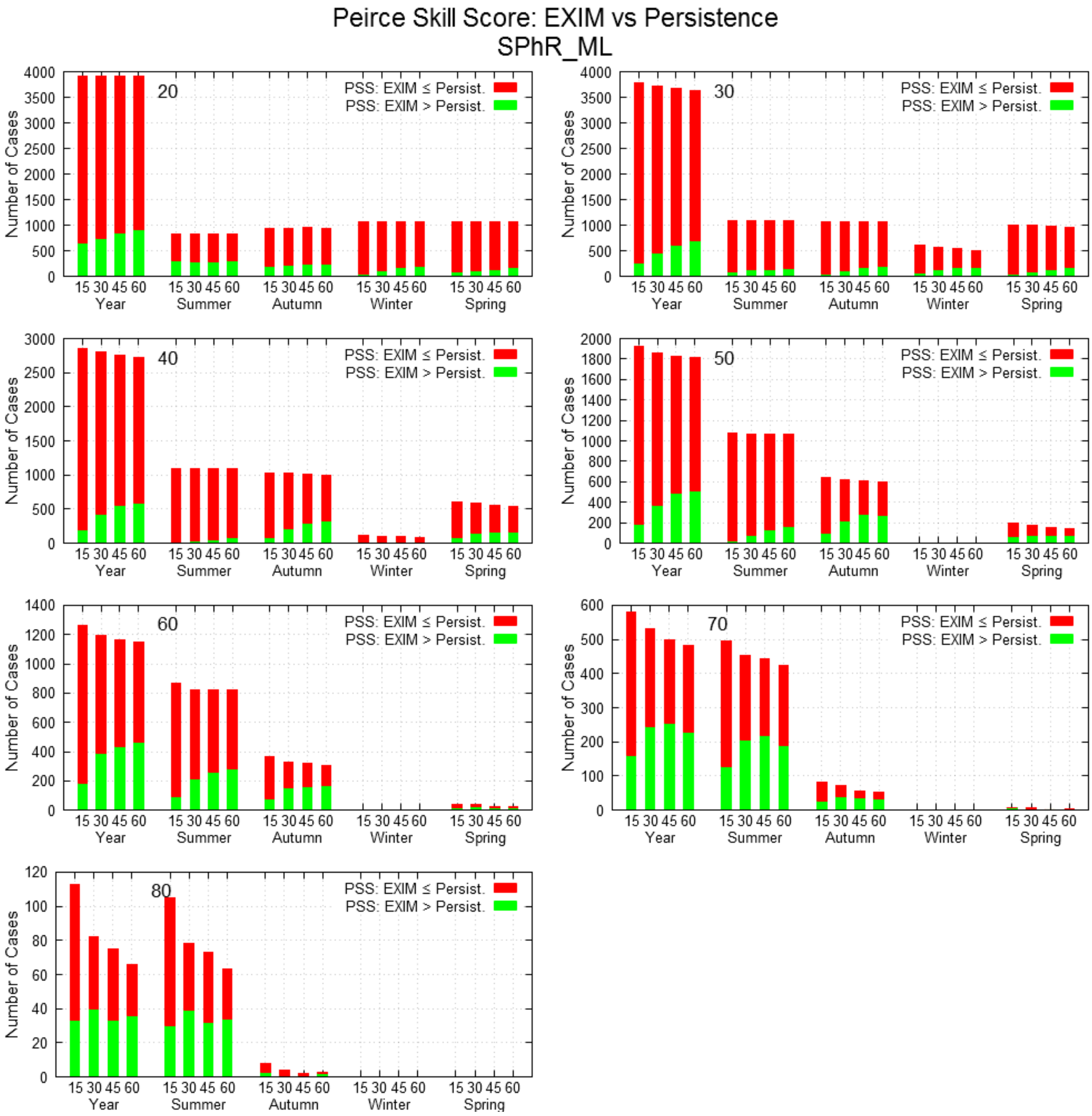

Figure 64: *As in Figure 46 but for SPhR_ML Product.*

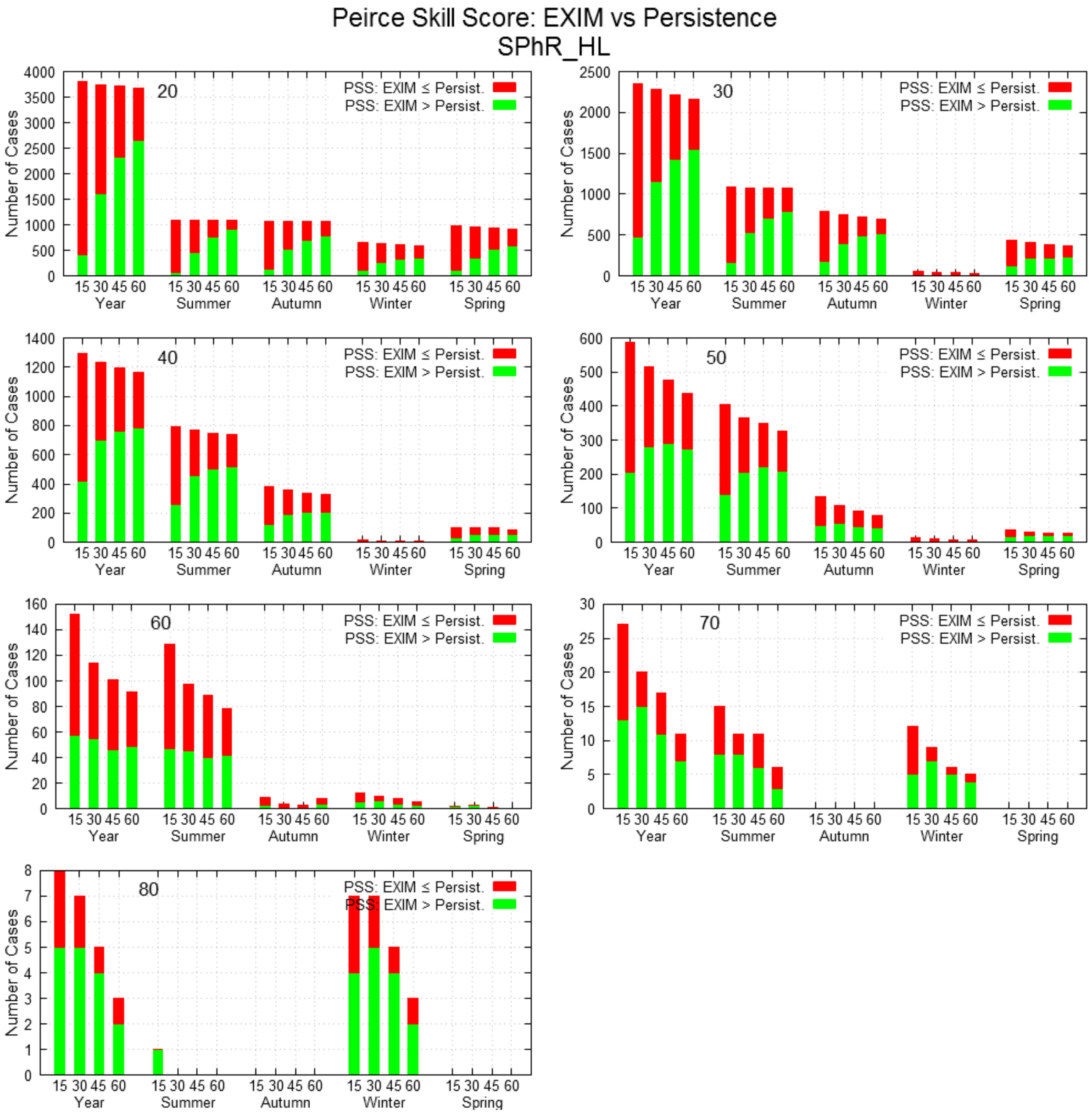

Figure 65: *As in Figure 46 but for SPhR_HL Product.*

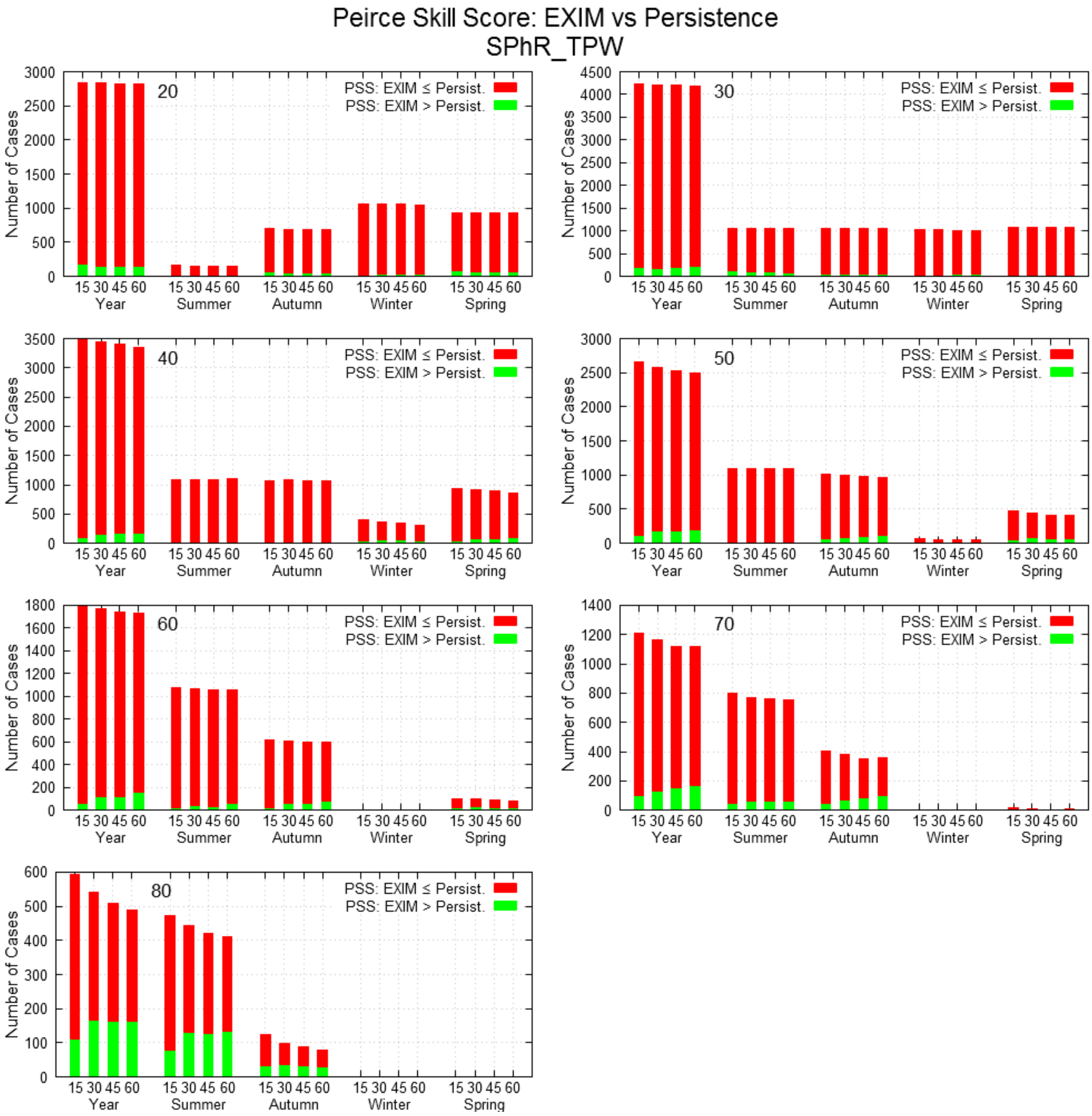

Figure 66: As in Figure 46 but for TPW Product.

A.11 SPhR: SEVIRI Physical Retrieval Products[4] – Stability Analysis

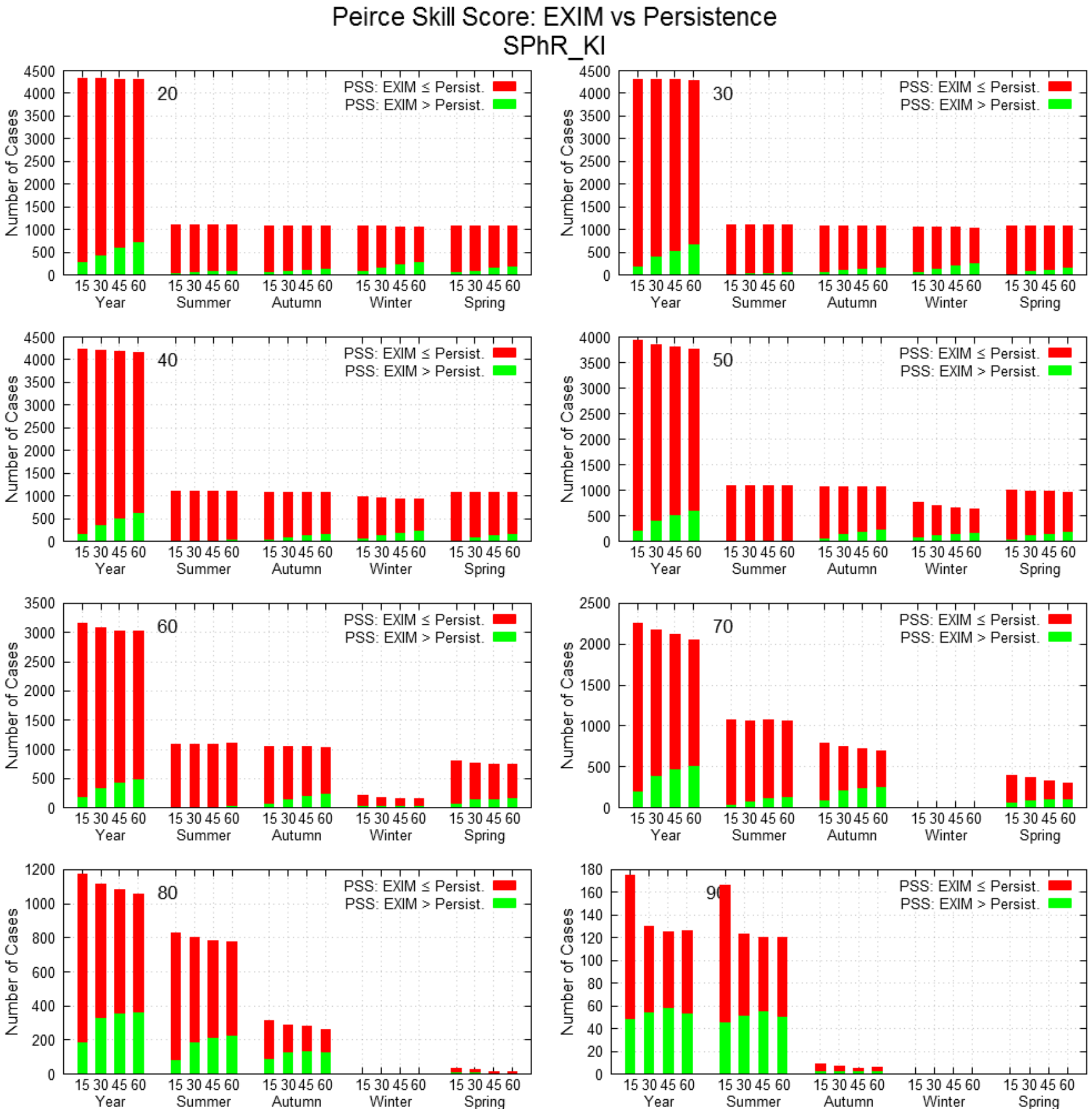

Figure 67: *As in Figure 46 but for SPhR_KI Product.*

[4] SPhR was the designation in v2013 (for which the validation was carried out). The product has been renamed in version NWC/GEO v2016 as iSHAI (Imaging Satellite Humidity and Instability).

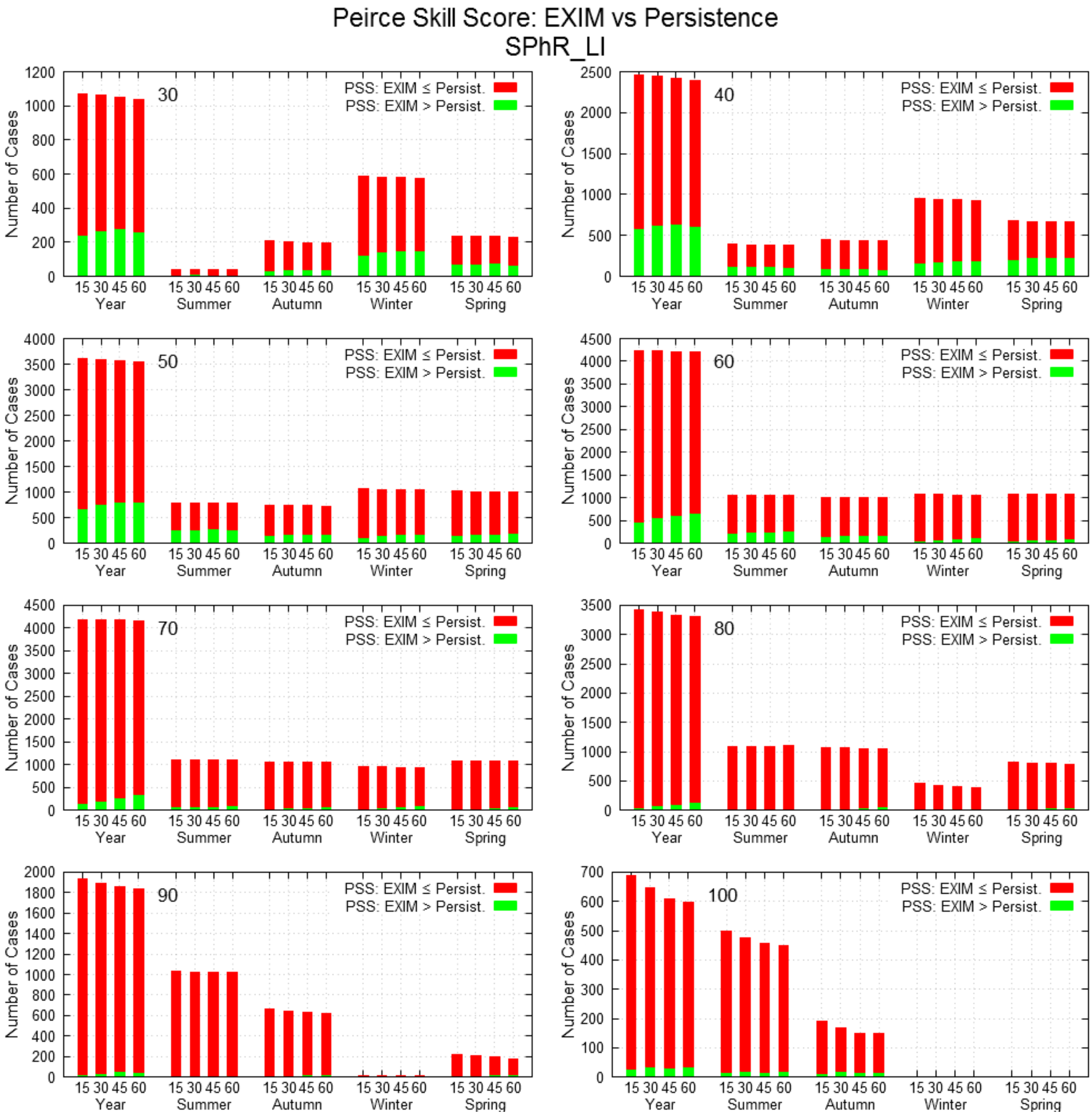

Figure 68: *As in Figure 46 but for SPhR_LI Product.*

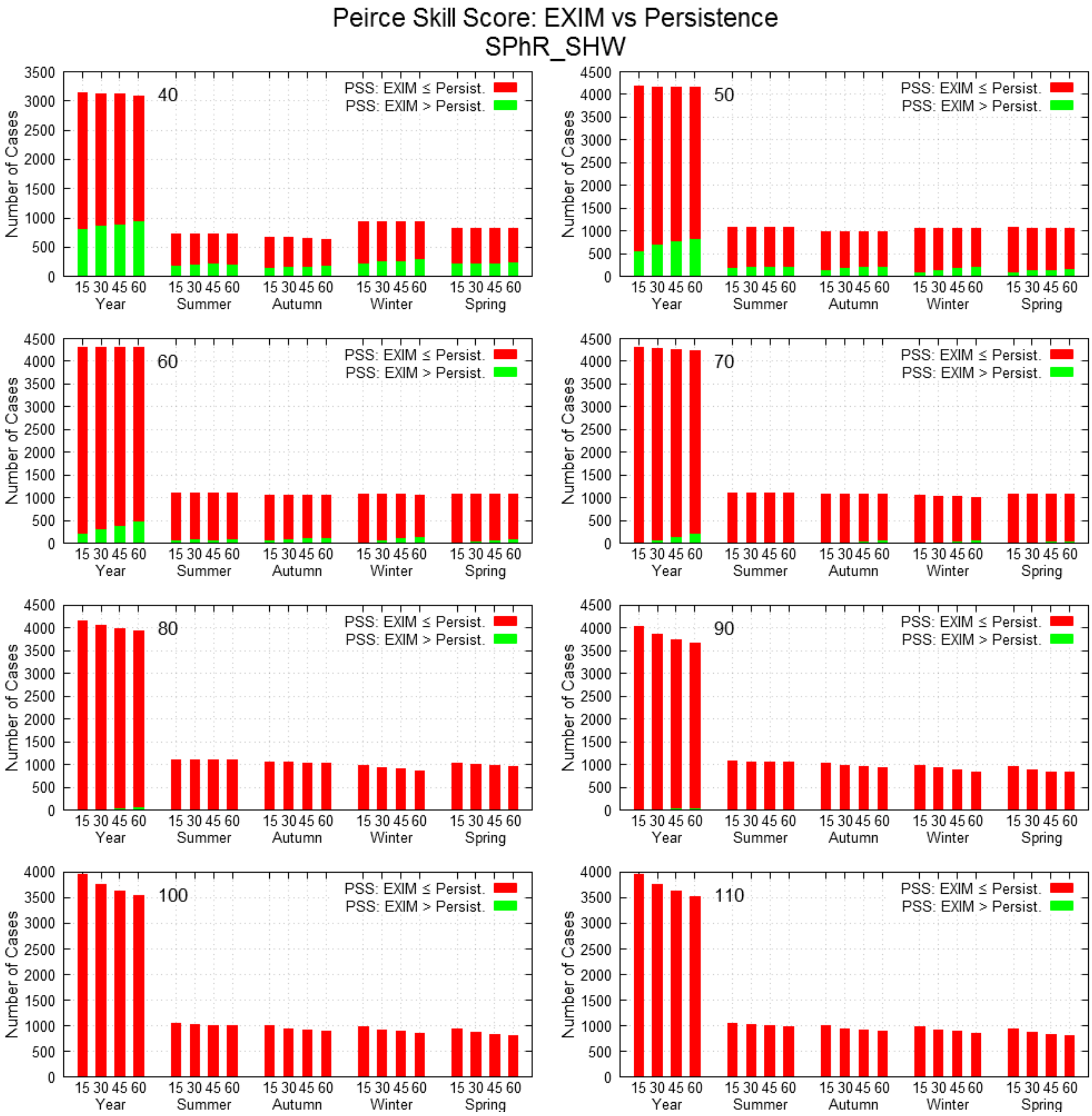

Figure 69: *As in Figure 46 but for SPhR_SHW Product.*

B. ANNEX: NOTES ADDED IN MAY 2017 FOR RELEASE NWC/GEO 2016, EXIM PATCH

The EXIM software underwent a few changes between the evaluations described in the main body of the text and its release in a patch to NWC/GEO v2016:

- In the first version, not only the clouds, but also cloudfree areas in VIS and IR imagery were displaced with the motion vectors. This led to moving coastlines, for example; the implemented remedy is to mask those areas using the NWC/GEO cloud mask.
- A single displacement field was used for all satellite channels and NWC/GEO products, mixing vectors derived from all channels and being assigned to any height. It was agreed later after some expert discussions to concentrate on extrapolation of the features most prominently reflected in the individual channels, and in turn to discard displacement vectors stemming from other layers. Hence, IR imagery is now extrapolated with high-level IR/VIS vectors (high-level means: < 400 hPa; the vectors are derived from VIS 0.6 and 0.8, HRVIS, IR 10.8 and 12.0; for GOES-N, just VIS 0.7 and IR 10.7 are available), VIS imagery with low-level IR/VIS AMVs (> 700 hPa). The extrapolated NWCSAF products are generally driven to a higher degree by the higher clouds, so the trajectories are the same as for IR imagery. The water vapour absorption bands have different characteristics which eventually allows to avoid any thresholding, i.e. WV6.2 is extrapolated with all WV6.2 that are generated by the HrW module (and analogously for WV7.3).

The obvious question is what the practical impact of these changes on the results presented in the main body of this validation report is. Unfortunately, the time to collect material from the modified EXIM software in operations was very short (from local implementation of NWC/GEO 2016 end-of-February until end-of-April 2017, and some heterogeneities that originated from experimenting with the new software package also impacted the skill score comparisons; these heterogeneities are listed below where applicable), so it has to be borne in mind that the following results should be considered with the due caution. Nonetheless, when comparing PSS for extrapolation with the PSS for persistence, the results for IR10.8 are:

Threshold	Percentage of where EXIM excels persistence (the four figures are for +15, +30, +45, +60 minutes lead time, respectively)
230 K	94.8 – 93.8 – 93.1 – 90.2
240 K	96.1 – 94.6 – 93.6 – 90.8
250 K	95.3 – 91.4 – 89.2 – 84.3
260 K	88.0 – 83.4 – 75.6 – 69.6

And for IR13.4:

Threshold	Percentage of where EXIM excels persistence (the four figures are for +15, +30, +45, +60 minutes lead time, respectively)
230 K	96.8 – 95.0 – 93.2 – 91.8
240 K	97.5 – 95.9 – 94.4 – 90.5
250 K	91.2 – 85.2 – 75.3 – 65.3

The usage of the mentioned cloud masking option was not consistent in the covered period, which makes statistics for the warmest thresholds questionable and they are withheld here. Still, the figures for the other temperature thresholds suggest that the superiority of EXIM is still there but being less pronounced than in the winter/spring columns of **Figure 6** (IR10.8) and **Figure 8** (IR13.4).

With respect to NWC/GEO products that are extrapolated with the same trajectories as the IR channels, some technical inconsistencies between v2013 and v2016 were detected (e.g. the cloud type product has one additional category, the CTTH product has a finer resolution in terms of metres; the cloud products were run in the local operational suite with different segment sizes in order to finally arrive at the one-pixel segments considered imperative for a fair EXIM evaluation, avoiding the blocky structures discussed in section 2.8). Eventually, only the figures for the CRR product were judged comparable enough with the statistics shown for v2013 (in section 2.9):

Threshold	Percentage of where EXIM excels persistence (the four figures are for +15, +30, +45, +60 minutes lead time, respectively)
1 mm	85.4 – 77.6 – 69.0 – 62.5
2 mm	70.6 – 64.5 – 58.8 – 51.5
3 mm	63.0 – 57.1 – 52.1 – 55.0

For higher precipitation values, the scores became erratic due to the small number of pixels with such values in the months of March and April. If compared with **Figure 33** (the rain amounts given in the table correspond to classes 1, 2, 3 of that figure), we obtain the same impression as for the IR channels: better than persistence, but the scores are not as favourable (for EXIM) as before. Statements on the possible root cause(s) are fairly speculative at this point:

- There is one methodological issue discovered during the present investigation: we retained cases with very few pixels above a threshold and then compared irrelevant and random skill scores close to 0. They had the same weight in the summary statistics as meaningful comparisons. The number of such cases may increase - pushing the score towards 50/50 - if many pixels on one side of the threshold are filtered by the cloud mask. As lesson learned, the significance of the skill scores should be better addressed in future.
- An obvious candidate for explaining deterioration is the reduced number of vectors describing the extrapolation field compared to the previous EXIM version (the WV vectors and all vectors below 400 hPa have been withdrawn), thus perhaps leading to an inferior description of the overall atmospheric movement. On the other hand, the statistics shown shortly for the water vapour channels indicate no deterioration for those channels though the number of vectors has also been reduced. Clearly, this issue requires more in-depth considerations (such as: which channels contribute most vectors, in which season etc.) before declared to be the governing factor to be improved upon.

Having already announced the results for the WV channels, they are for WV6.2:

Threshold	Percentage of where EXIM excels persistence (the four figures are for +15, +30, +45, +60 minutes lead time, respectively)
220 K	98.5 – 95.1 – 90.4 – 85.1
225 K	100.0 – 99.9 – 99.4 – 98.6
230 K	99.9 – 100.0 – 99.5 – 98.6
235 K	93.7 – 90.8 – 89.5 – 88.4

And for WV7.3:

Threshold	Percentage of where EXIM excels persistence (the four figures are for +15, +30, +45, +60 minutes lead time, respectively)
220 K	92.8 – 90.3 – 85.1 – 80.8
230 K	100.0 – 99.5 – 98.4 – 96.5
240 K	100.0 – 100.0 – 99.8 – 99.6
250 K	100.0 – 99.6 – 98.1 – 96.1

When these figures are compared with winter/spring statistics in **Figure 17** (WV6.2) and **Figure 18** (WV7.3), one cannot substantiate a change in the performance of the EXIM output with the new setting.

The near future will see an introduction of difference image computation (last forecast minus current observation) into the EXIM software. Originally planned to satisfy a user request on NRT quality monitoring, the EXIM developers shall use this by-product of the software to identify cases with non-random forecast errors (and to evaluate the effects of taken countermeasures). It is intended in the next round of validation efforts to specifically address such situations in order to better understand which screws to turn in EXIM (or the HrW input) to be able to again report the slightly higher performance figures we found for the precursor.