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

NWC/CDOP3/GEO/ZAMG/SCI/VR/EXIM, Issue 2.0.1

28 February 2022

Applicable to

GEO-EXIM-v2.1 (NWC-087)

Prepared by ZAMG

For quick navigation go to: Table of contents

REPORT SIGNATURE TABLE

Function	Name	Signature	Date
Prepared by	P. Schmederer (ZAMG)		*28 February 2022*
Reviewed by	A. Jann (ZAMG)		*28 February 2022*
Authorised by	P. Rípodas (AEMET) *SAFNWC Project Manager*		*28 February 2022*

DOCUMENT CHANGE RECORD

Version	Date	Pages	CHANGE(S)
2.0	*10 January 202228 February 2022*	46	Creation (Document with version number 1.0 applied to NWC/GEO v2018)
2.0.1	*28 February 2022*	46	Section 3.2.1 updated after DRR for GEO/v2021

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 (*EXIM*) Products (PGE16), for the NWC/GEO release 2021.

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
CMA	Cloud Mask
CMIC	Cloud Microphysics
CRR	Convective Rainfall Rate
CRRPh	Convective Rainfall Rate from Cloud Physical Properties
CSI	Critical Success Index
CT	Cloud Type
CTTH	Cloud Top Temperature and Height
ECMWF	European Centre for Medium-Range Weather Forecasts
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EXIM	Extrapolated Imagery
FAR	False Alarm Ratio
HRW	High-Resolution Winds
IFS	Integrated Forecasting System
IR	Infrared
MSG	Meteosat Second Generation
NWC	Nowcasting

NWP	Numerical Weather Prediction
PC	Precipitating Clouds
PCPh	Precipitating Clouds from Cloud Physical Properties
PGE	Product Generation Element
POFD	Probability of False Detection
POD	Probability of Detection
PSS	Peirce Skill Score
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]	Project Plan for the NWCSAF CDOP3 phase	NWC/CDOP3/SAF/AEMET/MGT/PP	1.6	01/12/21
[AD.2]	NWCSAF CDOP3 Project Plan Master Schedule	NWC/CDOP3/SAF/AEMET/MGT/PP/MasterSchedule	1.6	01/12/21
[AD.3]	Configuration Management Plan for the NWC SAF	NWC/CDOP3/SAF/AEMET/MGT/CMP	1.1	15/04/20
[AD.4]	System and Components Requirements Document for the NWC/GEO	NWC/CDOP3/GEO/AEMET/SW/SCRD	1.0	01/09/21
[AD.5]	Interface Control Document for Internal and External Interfaces of the NWC/GEO	NWC/CDOP3/GEO/AEMET/SW/ICD/1	2.0	01/09/21
[AD.6]	Interface Control Document for the NWCLIB of the NWC/GEO	NWC/CDOP3/GEO/AEMET/SW/ICD/2	2.0	01/09/21
[AD.7]	Data Output Format for the NWC/GEO	NWC/CDOP3/GEO/AEMET/SW/DOF	2.0	10/01/22
[AD.8]	Component Design Document for the NWCLIB of the NWC/GEO	NWC/CDOP2/GEO/AEMET/SW/ACDD/NWCLIB	2.0.1	31/07/18
[AD.9]	NWC SAF Product Requirements Document	NWC/CDOP3/GEO/AEMET/MGT/PRD	1.5	01/12/21
[AD.10]	User Manual for the Tools of the NWC/GEO	NWC/CDOP3/GEO/AEMET/SCI/UM/Tools	2.0	10/01/22

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/CDOP3/SAF/AEMET/MGT/GLO	1.0	20/10/20
[RD.2]	User Manual for the Extrapolated Imagery Processor of the NWC/GEO: Science Part	NWC/CDOP3/GEO/ZAMG/SCI/UM/EXIM	2.0	10/01/22
[RD.3]	Algorithm Theoretical Basis Document for the Extrapolated Imagery Processor of the NWC/GEO	NWC/CDOP3/GEO/ZAMG/SCI/ATBD/EXIM	1.0	01/09/21
[RD.4]	Scientific and Validation Report for the Extrapolated Imagery Processor of the NWC/GEO	NWC/CDOP2/GEO/ZAMG/SCI/VR/EXIM	1.0	22/05/17
[RD.5]	Algorithm Theoretical Basis Document for the Precipitation Product Processors of the NWC/GEO	NWC/CDOP3/GEO/AEMET/SCI/ATBD/Precipitation	1.0.1	29/10/21

Table 2: *List of Referenced Documents*

2. GENERAL ASPECTS OF THE VALIDATION APPROACH

2.1 VALIDATION OBJECTIVES

The validation report covers three main aspects that will be presented in the following.

2.1.1 Product evaluation of EXIM's applicability

A new sub-product ("*CTTH effectiv*", from the *CTTH* product) has been added to the list of products being forecasted by *EXIM* and two of the input products (*PCPh*, *CRRPh*) have undergone substantial changes in the underlying algorithms. Those three sub-products of *EXIM* are evaluated to verify the superiority against persistence forecasts. According to the NWCSAF Product Requirements Table (PRT) ([AD.9]), the threshold accuracy of *EXIM* is described by "*on average better than persistence forecast*" and the target accuracy is "*always better than persistence forecast*".

All three products are evaluated using the current default model version of *EXIM*. The setup is a one-layer scheme with lower and upper boundaries as defined in Chapter 2.1.2 , thereunder named as "control setup". Following a user request, cloud top temperature and height's sub-product "effective cloudiness" (*CTTH effectiv*) has been added to the list of products being extrapolated by *EXIM*. Effective cloudiness is defined as fraction of the field of view covered by clouds (cloud amount) multiplied by the cloud emissivity in the 10.8 µm window channel. The values of *CTTH effectiv* range from 0 to 1, with 1 meaning "thick clouds" and values below 1 signalling semi-transparent clouds. Like all new products, which are added to *EXIM*, *CTTH effectiv* will be evaluated against persistence to ensure that *EXIM*'s extrapolation method can beneficially be applied to this product and results in a gain of forecast skill compared to pure persistence (see [RD.4]).

Also, the NWCSAF products Precipitating Clouds from Cloud Physical Properties (*PCPh)* and Convective Rainfall Rate from Cloud Physical Properties (*CRRPh*) underwent some improvements of their algorithms (see [RD.5]). This validation report will evaluate whether the *EXIM* forecasts for the two products still outperform persistence, i.e. they fulfil the threshold accuracy (see SAFNWC Product Requirements Table [AD.9]). *PCPh* provides an estimation on the probability of precipitation occurrence which is similar to *PC*. *PCPh* uses cloud physical properties such as cloud top microphysical properties, effective radius, and cloud optical thickness to derive the probability of precipitation occurrence. *CRRPh* estimates rain rates associated to convection. This Nowcasting tool uses cloud physical properties generated by *CMIC* such as cloud top phase, effective radius and cloud optical thickness. To learn more about the changes applied to the algorithm of *PCPh* and *CRRPh*, please refer to [RD.5].

2.1.2 Evaluation of CTTH Filter

With version v2021, a new feature has been implemented in the *EXIM* software. There is now the possibility of using the so-called CTTH filter. With the CTTH filter in use, pixels will be extrapolated only by AMVs stemming from the same layer. The user can choose up to two layers. The idea is being able to distinguish between lower and upper weather phenomena or to pick a specific layer of interest and exclude AMVs stemming from other heights.

This validation report will evaluate the performance of the filter for the products (*CMA*, *CMIC*, *CT*, *PC*, *PCPh*, *CRR*, *CRRPh*, *CTTH*) as well as the satellite channels (*IR108*, *IR38*, *VIS08* and *VIS06*) and result in a recommendation if and to what extent the filter shall be used. Only two IR channels have been picked to save calculation capacity and since the same behaviour can be expected for all of them (as indicated in previous validation reports (VRs)).

The 5 configurations for this evaluation are:

- "control setup": pixels of ***all*** heights are extrapolated with vectors from the heights 900 – 100 hPa
- "Low, no filter": pixels of ***all*** heights are extrapolated with vectors from the **low** layer (900 – 500 hPa)
- "Low, with filter": pixels from the **low** layer (900 – 500 hPa) are extrapolated with vectors from the **low** layer (900 – 500 hPa)
- "High, no filter": pixels from ***all*** heights are extrapolated with vectors from the **high** layer (500 – 100 hPa)
- "High, with filter": pixels form the **high** layer (500 – 100 hPa) are extrapolated with vectors from the **high** layer (500 – 100 hPa)
- "2-layer, with filter": pixels of **low** heights (900 – 501 hPa) are extrapolated with vectors from **low** heights (900 – 501 hPa) and pixels from the **high** layer (500 – 100 hPa) are extrapolated with vectors from the **high** layer (500 – 100 hPa)

2.1.3 Evaluation of including WV vectors

The question arose whether water vapour (WV) atmospheric motion vectors (AMVs) should be included to the set of used AMVs for all products. Currently, WV AMVs are only used for the extrapolation of the WV channels. All other products and channels by default use AMVs from channels: IR108, IR120, VIS06, VIS08, HRVIS.

The configurations for this evaluation are:

- "control, without WV AMVs": same as "control setup" from Chapter 2.1.2; AMVs from: IR108, IR120, VIS06, VIS08, HRVIS
- "setup with WV AMVs": similar to "control, without WV AMVs" with **additional AMVs** from: WV062, WV073

2.2 DATA

EXIM version 2.1 of the NWCSAF/GEO software v2021 has been subject of this evaluation. The validation dataset covers the period 1st April to 31st August 2021 for all products, except *VIS08*, *IR108* and *VIS06* which only start on 23, 23 and 29 April 2021, respectively. The set of extrapolated SEVIRI channels and NWCSAF products are listed in Table 3. As marked in Table 3 and described in Chapter 2.1.1, only three products have been validated against persistence. Due to their new algorithms or being recently added to the list of extrapolated products, their value of being extrapolated by *EXIM* needs to be proved (for the other products, this has been verified in earlier VRs). However, for all products listed in the table this study evaluates firstly, whether WV AMVs should be included in the set of AMVs. Secondly, which model version is superior depending on the heights the AMVs and the extrapolated pixels are from.

For the sake of completeness, a 2-layer scheme with CTTH filter has been added on 6 June 2021 to the range of setups. Since those scores stem from a shorter time period, their results shall be handled with care when comparing them with the other setups.

Scores have been logged every hour (see the range of scores in Chapter 2.3) and for forecasts with lead times of +15, +30, +45 and +60 min. The region of the study is Europe and is shown in Figure 2.1.

Due to their nature, visible channels have only been considered at daytime between 06 and 18 UTC.

Forecasts from the integrated forecasting system (IFS) from the European Centre for Medium Range Weather Forecasts (ECMWF) have been used as numerical weather prediction (NWP) input.

The versions of each product are listed in Table 3, high resolution winds (HRW) have been calculated with version 7.0.

<table>
<tr><th colspan="2" rowspan="2">Product (version)</th><th rowspan="2">Abbreviation</th><th rowspan="2">Details</th><th colspan="3">Evaluated aspects of chapter</th></tr>
<tr><th>2.1.1</th><th>2.1.2</th><th>2.1.3</th></tr>
<tr><td colspan="2">SEVIRI thermal infrared</td><td>IR3.9, IR10.8</td><td>3.9 µm, 10.8 µm</td><td></td><td>x</td><td>x</td></tr>
<tr><td colspan="2">SEVIRI visible</td><td>VIS0.6, VIS0.8</td><td>0.6 µm, 0.8 µm</td><td></td><td>x</td><td>x</td></tr>
<tr><td colspan="2">Convective Rainfall Rate (v5.0)</td><td>CRR</td><td></td><td></td><td>x</td><td>x</td></tr>
<tr><td colspan="2">Convective Rainfall Rate from Cloud Physical Properties (v3.0)</td><td>CRRPh</td><td></td><td>x</td><td>x</td><td>x</td></tr>
<tr><td colspan="2">Cloud Mask (v5.0)</td><td>CMA</td><td></td><td></td><td>x</td><td>x</td></tr>
<tr><td colspan="2">Cloud Type (v4.0)</td><td>CT</td><td></td><td></td><td>x</td><td>x</td></tr>
<tr><td rowspan="2">Cloud Top Temperature and Height (v4.0)</td><td>altitude</td><td>CTTH alti</td><td></td><td></td><td>x</td><td>x</td></tr>
<tr><td>effective cloudiness</td><td>CTTH effectiv</td><td></td><td>x</td><td>x</td><td>x</td></tr>
<tr><td colspan="2">Precipitating Clouds (v2.0)</td><td>PC</td><td></td><td></td><td>x</td><td>x</td></tr>
<tr><td colspan="2">Precipitating Clouds from Cloud Physical Properties (v3.0)</td><td>PCPh</td><td></td><td>x</td><td>x</td><td>x</td></tr>
<tr><td colspan="2">Cloud Microphysics (v2.0)</td><td>CMIC phase</td><td></td><td></td><td>x</td><td>x</td></tr>
</table>

Table 3*: SEVIRI channels and NWCSAF products evaluated in this validation report.*

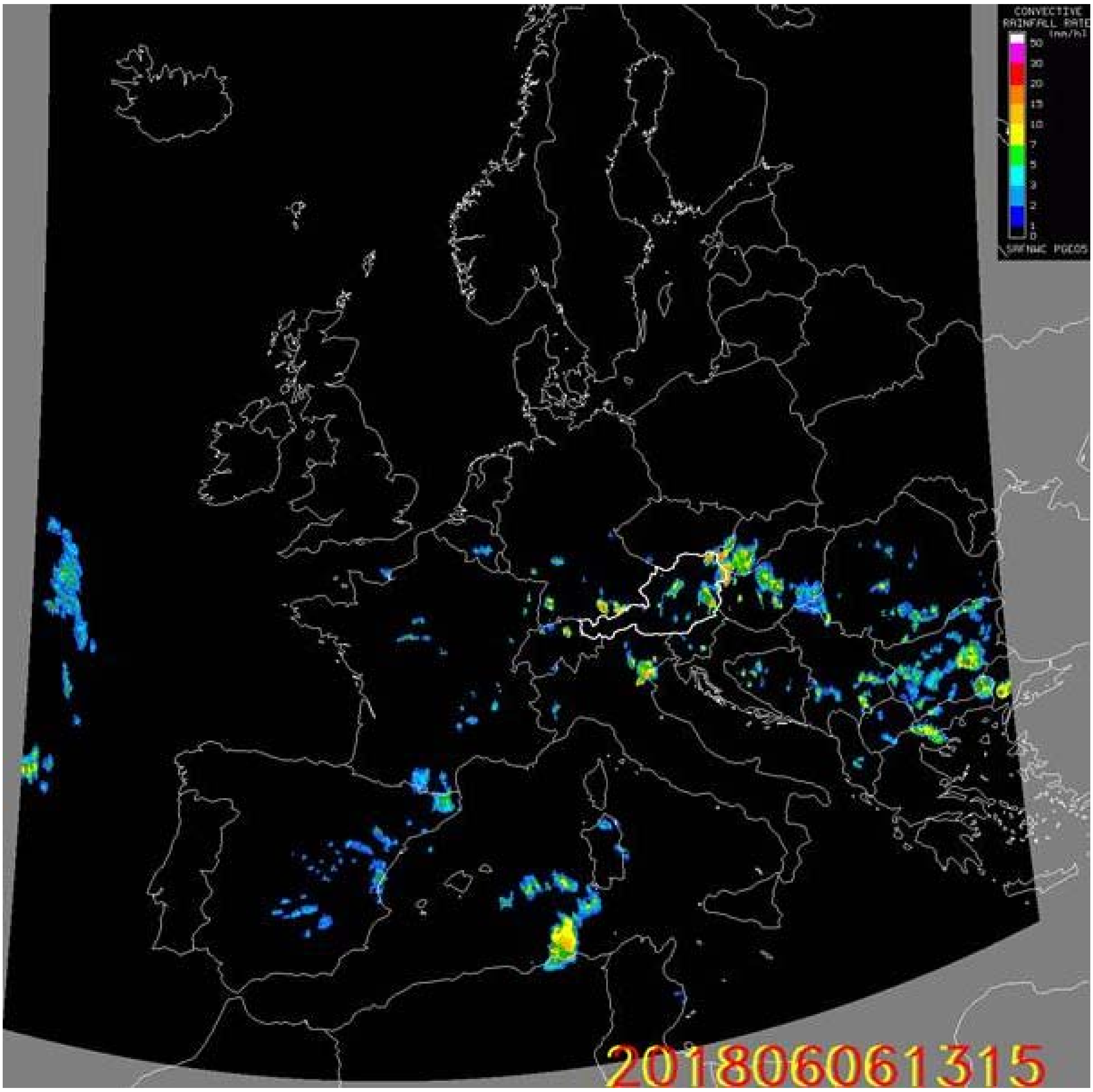


Figure 2.1 Geographical area over which the analyses have been performed.

2.3 METHODOLOGY

Forecasts (meaning forecasts of *EXIM* and persistence “forecasts”) have been validated against satellite images and derived products at verification time. For the underlying analysis, dichotomous scores such as POD, POFD, FAR, CSI and PSS (find definitions of each score below) were used. Dichotomous scores separate the data, in this case pixel-wise, in “yes, the event will happen” and “no, the event won’t happen”. For most of the products/ satellite images thresholds have been specified by separating “yes” and “no”, in the sense of “the value is greater than/equal to the threshold” and “the value is smaller than the threshold”, respectively. Multi-category products were classified similarly for each category and a multi-category skill score was computed additionally in a slightly different manner, as described below.

The scores of the *EXIM* forecasts have been compared either against a control setup or against persistence. As stated above, threshold accuracy is being "*on average better than persistence*.

To calculate dichotomous scores, one starts with a contingency table (**Table 4**) that shows the frequency of forecasts and occurrences in the domain and their joint distribution of hits (a), false alarms (b), misses (c) and correct negatives (d). Such a contingency table was produced for every lead time (+15, +30, +45, +60 min) and resulted in a set of scores for each lead time and time step averaging over the whole domain.

The used scores are defined as described in the following. For a more comprehensive description see https://www.cawcr.gov.au/projects/verification/ or Jolliffe and Stephenson[1] (2012).

	observed			
		yes	**no**	total
forecast	**yes**	a hits	b false alarms	a + b forecast yes
	no	c misses	d correct negatives	d + d forecast no
	total	a + c observed yes	b + d observed no	n = a + b + c + d total

***Table 4**: Contingency table showing frequency of "yes" and "no" forecasts and occurrences.*

2.3.1 POD

The probability of detection (POD) answers the question what fraction of observed "yes" events was correctly forecasted. This score is good for rare events but ignores false alarms.

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

2.3.2 FAR

The false alarm ratio (FAR) answers the question what fraction of predicted "yes" events actually did not occur. This score ignores misses and is good for rare events.

It is defined as:

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

[1] Jolliffe, I.T., and D.B. Stephenson, 2012: Forecast Verification: A Practitioner's Guide in Atmospheric Science. 2nd Edition. Wiley and Sons Ltd, 274 pp.

2.3.3 POFD

The probability of false detection (POFD), also called false alarm rate (F) answers the question what fraction of observed “no” events were incorrectly forecasted as “yes”. This score ignores misses. Note that it can be artificially improved by issuing fewer “yes” forecasts to reduce the number of false alarms.

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

2.3.4 CSI

The critical success index (CSI), also denoted threat score (TS) answers the question how well forecasted “yes” events correspond to observed “yes” events. It can be thought of accuracy when correct negatives have been removed.

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

2.3.5 PSS

The Peirce Skill Score (PSS), also known as true skill statistic (TSS) or Hanssen and Kuipers discriminant (HK), answers the question how well the forecast separates the “yes” events from the “no” events. PSS is a measure of skill obtained by the difference between POD and POFD and is defined as:

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

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

	Observed category					
Forecasted category	i, j	1	2	…	K	Total
	1	n(F_1, O_1)	n(F_1, O_2)	…	n(F_1, O_K)	N(F_1)
	2	n(F_2, O_1)	n(F_2, O_2)	…	n(F_2, O_K)	N(F_2)
	…	…	…	…	…	…
	K	n(F_K, O_1)	n(F_K, O_2)	…	n(F_K, O_K)	N(F_K)
	Total	N(O_1)	N(O_2)	…	N(O_K)	

***Table 5**: Multi-category contingency table showing frequency of forecasts and occurrences in various bins.*

Two products are multi-category forecasts and therefore need to be treated slightly different. The products are:

- Cloud Type (*CT*), with 15 different cloud types;
- Cloud Microphysics' sub-product cloud phase (*CMIC phase*), with the five categories "liquid", "ice", "mixed", "cloud-free", and "un-defined".

For those, a multi-categorical variant of the PSS was used (PSS_{mc}). This method also starts with a contingency table (Table 5) showing the frequency of forecasts and occurrences of each bin. PSS_{mc} is defined as:

$$PSS_{mc} = \frac{\frac{1}{N}\sum_{i=1}^{K} n(F_i O_i) - \frac{1}{N^2}\sum_{i=1}^{K} N(F_i)N(O_i)}{1 - \frac{1}{N^2}\sum_{i=1}^{K}(N(O_i))^2} \qquad \text{Range } [-1,1].$$

3. RESULTS

Results for all three evaluation aspects,

- Comparison against persistence,
- Evaluation of CTTH Filter,
- Evaluation of including WV vectors,

are discussed in this chapter.

3.1 Comparison against persistence

Products newly added to *EXIM* and products that underwent a change in their algorithm are validated in the following.

Evaluations presented in this chapter are evaluated against persistence. Persistence means the initial state at lead time +0 minutes, which will be kept and with which the forecasts of *EXIM* of all lead times will be compared with. Requirement for *EXIM's* forecasts is to gain better scores than persistence.

3.1.1 CTTH effectiv: Cloud Top Temperature and Height – effective cloudiness

The product *CTTH effectiv* has been added in NWC/GEO v2021 to the range of products being forecasted by *EXIM*. *CTTH effectiv* provides effective cloudiness values ranging from 0 to 1; the thresholds for which dichotomous scores were derived are: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9.

EXIM's forecast for *CTTH effectiv* improves the scores compared to persistence, as illustrated in Figure 3.1. POD is improved in forecasts of *EXIM* compared to persistence in more than 95 % of all cases for an effective cloudiness greater 0.3. This improvement is observed at all lead times. FAR is reduced by *EXIM* in the majority of the cases (more than 50 %), like POD for effective cloudiness thresholds greater 0.3. This is valid for all lead times, except lead time 60 where the improvement of forecasts of *EXIM* only starts with threshold greater 0.4. The percentage of cases, where FAR is reduced by *EXIM*, rises up to above 95 % with increasing effective cloudiness. Similar to FAR, skill of POFD is improved in the majority of all cases for transparency thresholds greater 0.3 and lead times 15 and 30. The greater the chosen threshold of effective cloudiness, the greater is the percentage of cases where forecasts of *EXIM* improve POFD compared to persistence. Percentages of improving cases rise up to 95 %. The improvement of all scores is stronger pronounced for shorter lead times than it is for greater lead times.

Qualitatively, the biggest improvements achieved by forecasts of *EXIM* are a reduction of POFD and an increase of PSS for transparencies greater 0.3 (Figure 3.2). The improvements of POFD and PSS reach up to a reduction and an increase, respectively, by 0.1. The improvement is less pronounced with decreasing effective cloudiness and increasing lead time. Changes of POD and FAR are of an order of magnitude smaller, ranging from just above 0 to 0.05.

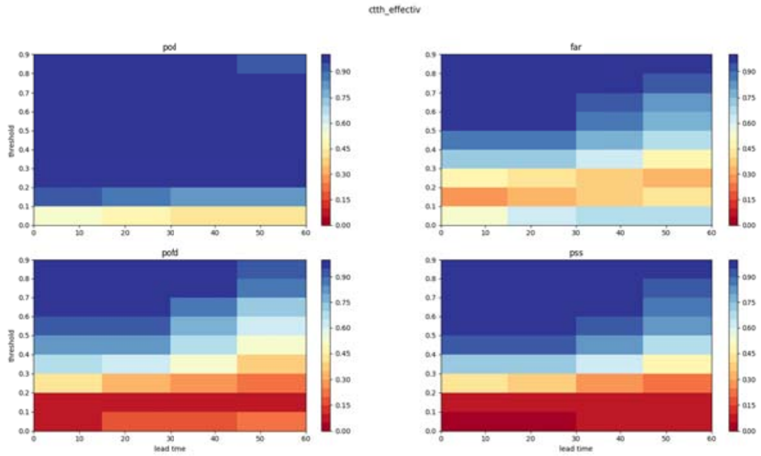


Figure 3.1 Frequency of cases where scores of the forecasts of CTTH effectiv are greater than their counterparts for persistence. Lead time vs. threshold categories. Top left: POD, top right: FAR, bottom left: POFD, bottom right: PSS. Blue values show an improvement, red a degradation.

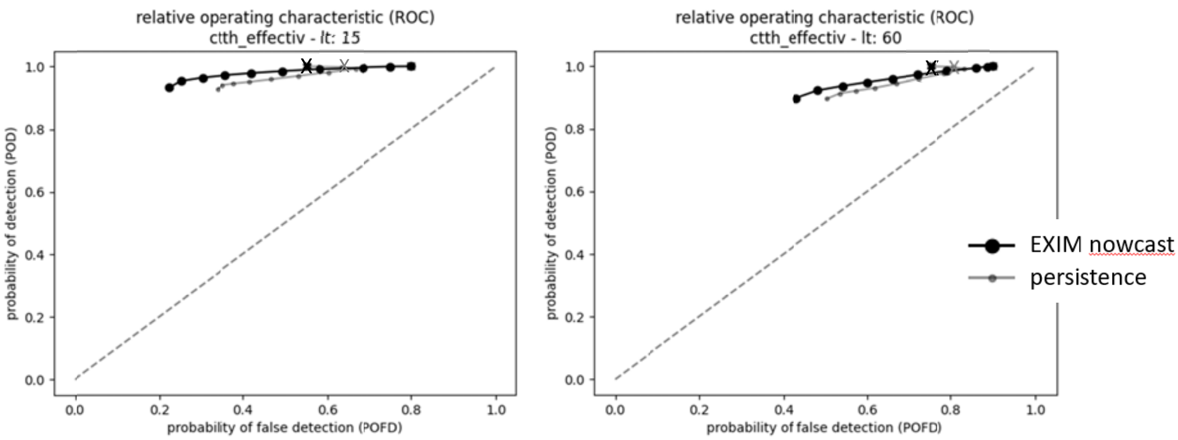


Figure 3.2 Relative operating characteristic (ROC) curve for CTTH effectiv. POFD vs POD. For details regarding the interpretation of this plot, please see Appendix 6.1. Threshold range: 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.8, 0.9. Each threshold 0.1 is marked with X for orientation.

3.1.2 PCPh: Precipitating Clouds from Cloud Physical Properties

PCPh is one of the two products that in NWC/GEO v2021 underwent a substantial change in algorithm. Due to the nature of the product, only daytime (between 06 and 18 UTC) has been considered. *PCPh* provides precipitation probabilities given in percentages and range from 11 % to 81 % with a step width of 10.

Forecasts of *EXIM* for *PCPh* improve the scores compared to pure persistence, as depicted in Figure 3.3. With a focus of precipitation probabilities greater than 11 %, *EXIM* achieves better scores than persistence in more than 95 % of all cases for both, POD and FAR. Above threshold 81 %, the number of cases in which *EXIM* outperforms persistence is slightly lower for FAR and significantly lower for POD. The percentage of improving cases decreases at probability threshold 81 % to 75 % and with increasing lead times down to 50 %, respectively. One has to keep in mind that "yes" events decrease with increasing thresholds, so only very few counts are left at threshold 81 %, and actual hits may occur at random.

Qualitatively, POD's and FAR's skill improves in forecasts of EXIM compared to persistence by about 0.1 (Figure 3.4) with the improvement being slightly lower for low *PCPh* probabilities, and slightly higher for greater *PCPh* probabilities. Forecasts of *EXIM* have a small bias towards too many "yes" events (threshold 11 %) and a slight bias towards too few "yes" events (thresholds greater and equal 71 %). Probabilities from 21 to 61 % are bias-free.

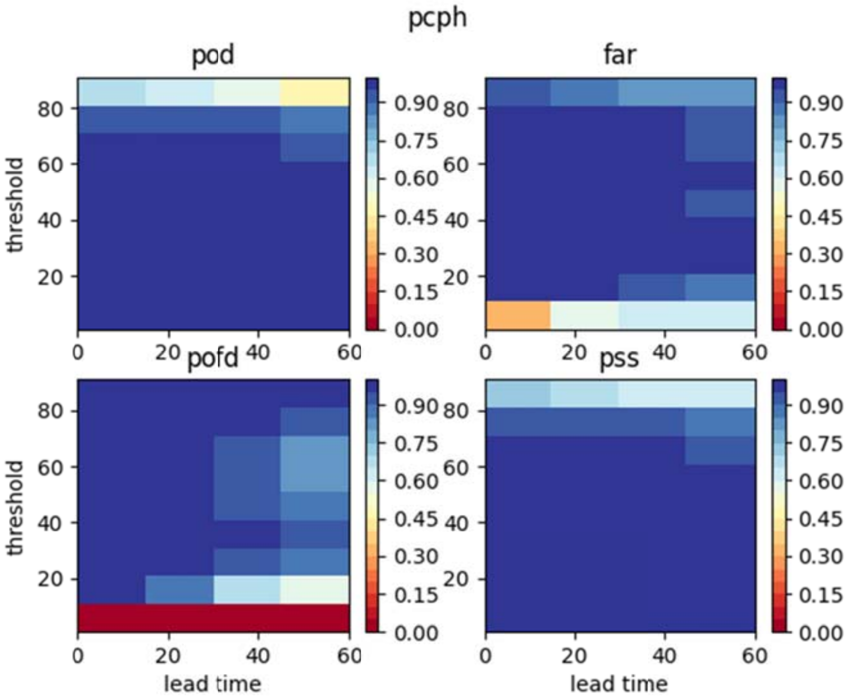


Figure 3.3 Frequency of cases where forecast skills for PCPh were greater than their counterparts for persistence. Lead time vs. threshold categories. Blue values show an improvement, yellow improvements and degradations of the same amount, red a degradation.

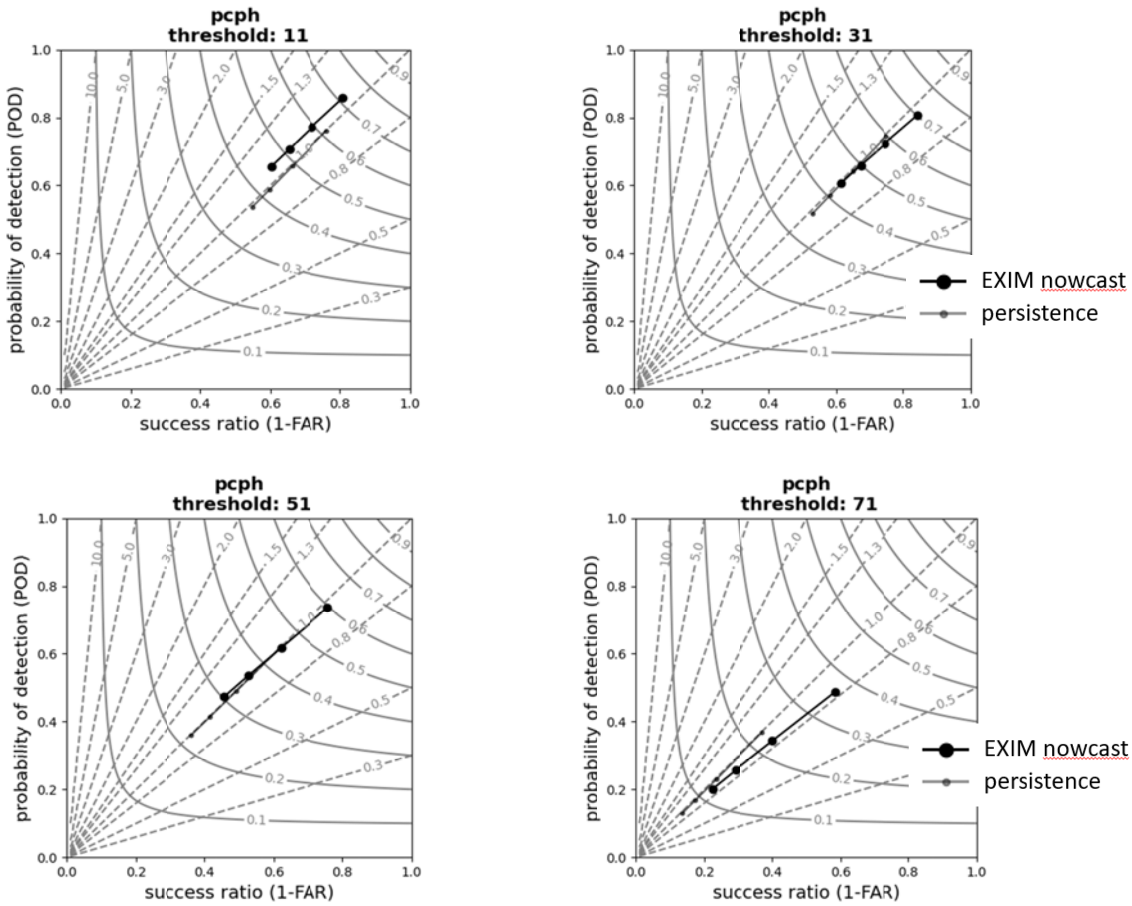


Figure 3.4 Performance diagram for PCPh. Top left: precipitation probability 11, top right: precipitation probability 31. Bottom left: precipitation probability 51, bottom right: precipitation probability 71. For details about the interpretation of these plots, please see Appendix 6.2.

3.1.3 CRRPh: Convective Rainfall Rate from Cloud Physical properties

CRRPh is the second of the two products that in NWC/GEO v2021 underwent a substantial change in algorithm. The product estimates rain rates from convective clouds through cloud top microphysical information such as cloud top effective radius and cloud optical thickness. The thresholds for which dichotomous scores were derived were chosen as: 1., 2., 3., 4., 5., 6., 7., 8., 9., 10., 11., 21. mm/h.

Forecasts of *EXIM* for *CRRPh* improve POD compared to persistence. In more than 95 % of all cases, forecasts of *EXIM* perform better than persistence at thresholds below 5 mm/h, as visible in Figure 3.5. The higher the thresholds and lead times, the less pronounced is the improvement - percentages of cases improving decrease with increasing lead times. At thresholds above 10 mm/h the amount of cases that improve POD drop below 50 % – persistence prevails value over the forecasts of *EXIM*. FAR improves in more than 95 % of all cases up to a threshold of 6 mm/h. With increasing thresholds and lead times, the improvement is getting less pronounced. The amount of the cases that improve drops to 80 % (see Figure 3.5).

Qualitatively, forecasts of *EXIM* for *CRRPh* add value to POD and FAR by an order of magnitude of about 0.1. Greater thresholds with short lead times improve FAR even slightly more, by about 0.15 (Figure 3.6). Forecasts of *EXIM* for *CRRPh* have a slight bias towards too few “yes” events at thresholds of 5 mm/h or greater and the lead times 15 and 30. However, the positive impact of POD and FAR is greater.

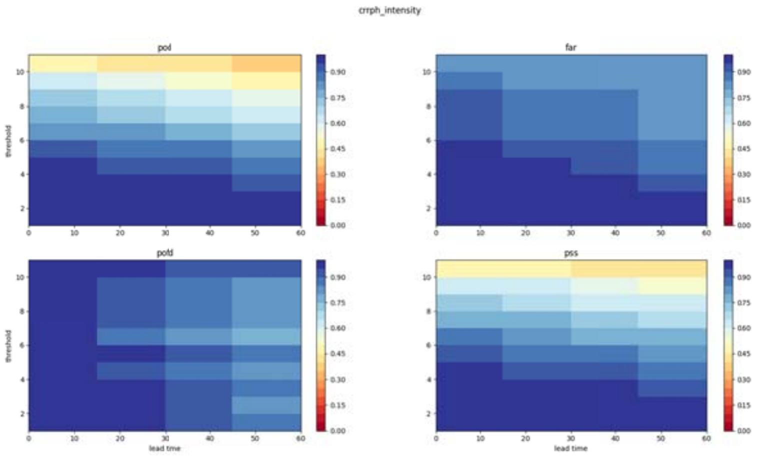


Figure 3.5 Frequency of cases where forecasts for CRRPh greater than its persistence for the four scores. Lead time vs. threshold categories. Blue values show an improvement, yellow values improvement and degradations of the same amount. Top left: POD, top right: FAR, bottom left: POFD, bottom right: PSS.

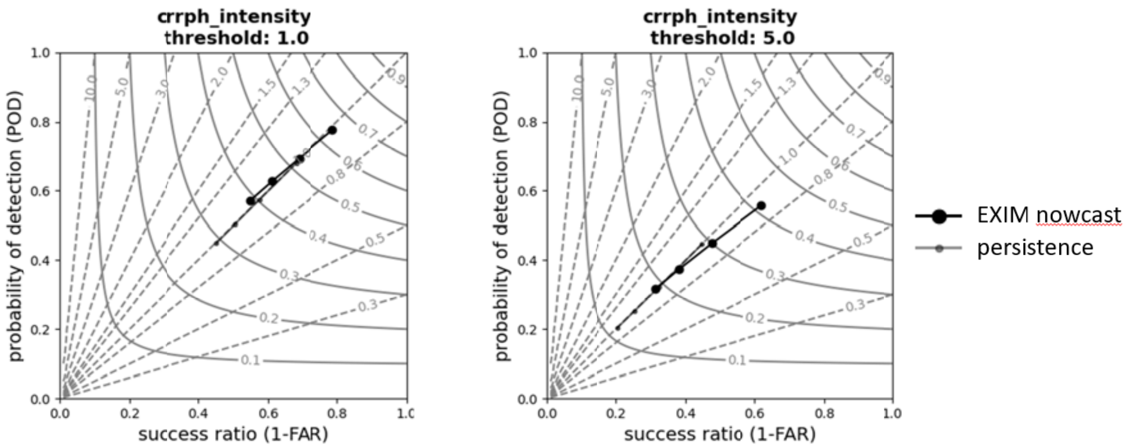


Figure 3.6 Performance diagram for CRRPh. Left: Precipitation intensity 1 mm/h, right: precipitation intensity 5 mm/h

3.2 Evaluation of CTTH Filter

The CTTH filter has been introduced to *EXIM* as a new feature in NWC/GEO v2021, allowing a filtering of pixels by height. Up to two layers can be defined by the user, separating e.g. a low-level from a high-level or selecting a specific layer of interest. With the filter in use, pixels will be extrapolated by *EXIM* using the closest set of atmospheric motion vectors (AMVs) stemming from the same layer as the pixels do. The idea is to avoid an extrapolation of pixels with AMVs stemming from a completely different height. For this validation report, two layers have been chosen, a lower level reaching from 900 to 500 hPa and an upper level reaching from 501 to 100 hPa. The following analysis will compare model setups as listed in Chapter 2.1.2. Each product will be evaluated separately in the following.

3.2.1 Thermal Channels: Infrared IR108

The infrared channel with wavelength *10.8 µm (IR108)* is evaluated in the range of 230 to 280 K brightness temperature (BT). *IR108* is strongly height dependent. Low BTs are rather correlating with high altitudes, while high BTs usually indicate low altitudes, however there are exceptions, such as semi-transparent clouds. With the filter in use, scores vary strongly with BTs, as visible in Figure 3.7. The low-level pixel extrapolation ("Low, with filter") starts to have better skill than the high-level pixel extrapolation ("High, with filter") for BTs greater 270 K (second highest BT). And for the high-level pixel extrapolation this is the other way around: Results for BTs smaller 250 K outperform low-level pixel extrapolation. As expected, for those BTs the two setups are rather not associated with, the setups perform way worse than all of the other setups. There is an overall decrease of scores with increasing lead times but pattern and behaviour are the same. POFD of "Low, with filter" has high values for low BT reaching up to 1 and POD of "High, with filter" has low values for high BT. Both setups with filter never perform best. This might be due to the fact that their sample size is smaller or caused by the chosen boundaries.

The two setups that extrapolate all pixels with either just low ("Low, no filter") or just high ("High, no filter") AMVs have a skill closest to the control setup's.

The forecasts by setup "2-layer, with filter" compete very well with control's forecasts for low BTs and slightly lose skill for greater BTs.

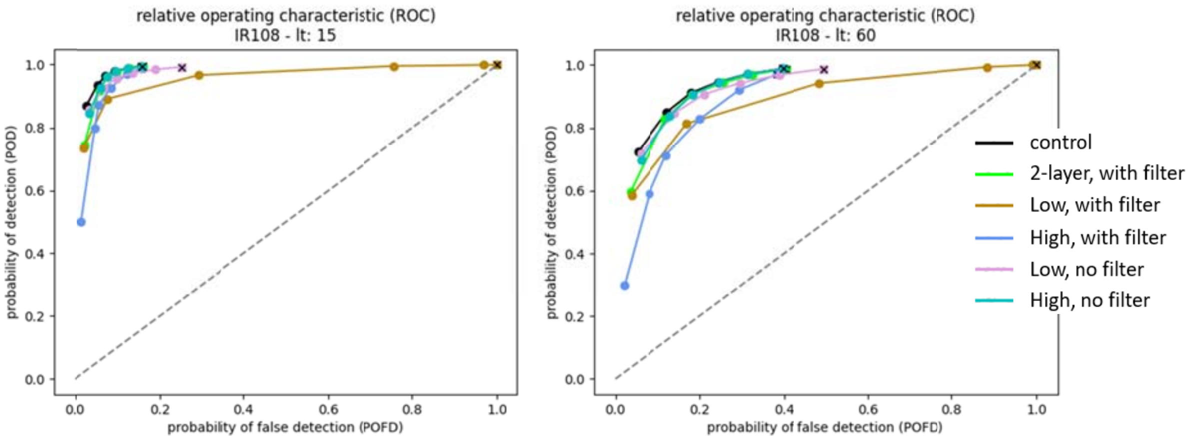


Figure 3.7 ROC curve for IR108. Left: lead time 15, right: lead time 60. Threshold range: 230, 240, 250, 260, 270, 280 K. Each 230 K is marked with X *for orientation.*

3.2.2 Thermal Channels: Infrared IR38

The infrared channel with wavelength *3.8 µm (IR38)* is evaluated in the range of 250 to 280 K BT. The behaviour is similar as for *IR108*. The height dependency is clearly visible when inspecting the scores in Figure 3.8. "Low, with filter" has huge POFD values for BTs smaller than 280 K. At daytime and for the lowest threshold, there is a general tendency of a too big POFD values in the forecasts of all setups (not shown), which is due to sun illumination not being negligible. When focussing on night time only, POFD is greatly reduced in the forecasts of all setups compared to an all-day-time analysis.

Like *IR108*, the two setups "Low, with filter" and "High, with filter" are performing worst for BT's associated with heights outside their layer. But also, within their associated layer, "Low, with filter" performs worst. At its best scenario, which is at lead time 15 and for low BTs, forecasts of "High, with filter" reach similar skill as the control setup.

Similar as for *IR108*, the control setup has highest scores for high BTs.

The forecasts of setup “2-layer, with filter” have similar high scores as control at low BTs and slightly loses skill at high BTs, which is most likely due to the chosen lower boundary of the layer.

IR38 is especially used at night to detect fog and very low clouds (less useful at daytime as the sun illumination is not negligible at this wavelength). Those might be missed by the setups with filter due to the lower boundary being 900 hPa, which should be kept in mind for a winter analysis.

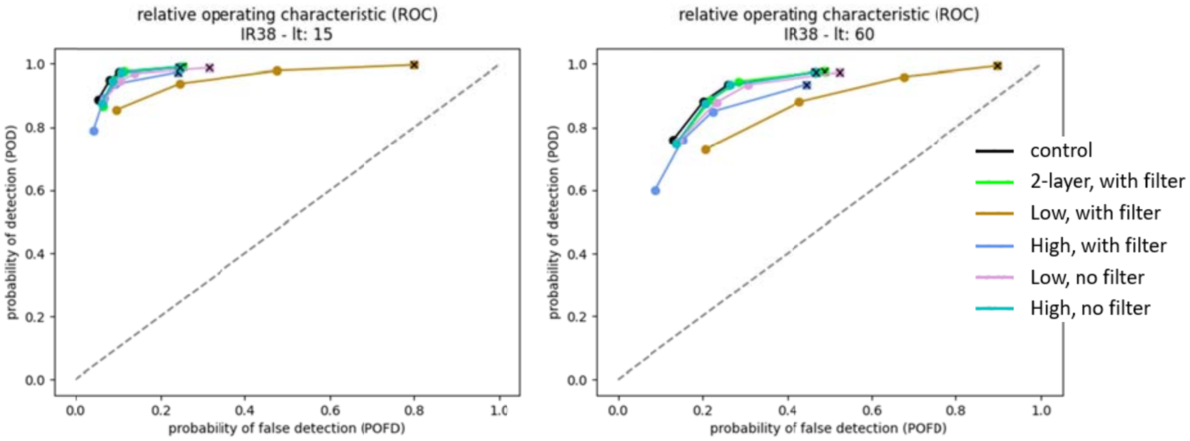


Figure 3.8 ROC curve for IR38. Left: lead time 15, right: lead time 60. Threshold range: 250, 260, 270, 280 K. Each 250 K is marked with X *for orientation.*

3.2.3 Visible Channels: VIS 06

The visible channel with wavelength *0.6 µm (VIS06)* is evaluated for reflectivity values of the value range 9, 15, 20, 30, 40, 50. Only day time, defined as 06 –18 UTC, has been considered due to the nature of this channel.

The ROC curve (see Figure 3.9) illustrates that using all AMVs for all pixels (control setup) is the superior option. The usage of the filter degrades POFD for low reflectivity values; The forecasts of “Low, with filter” degrade compared to control for POFD by up to 0.4 and the forecasts of “High, with filter” by 0.25. While setup “High, no filter” also slightly decreases POFD, forecasts of “Low, no filter” is of the same quality as the forecasts of the control setup. In the low reflectivity range (see left graph in Figure 3.9), POD is for all setups fairly the same. Also for low reflectivity range, there is a bias towards too many “yes”-events. This bias is slightly increased by the usage of the filter compared to the setups without the filter. And the bias increases for greater lead times.

The higher the reflectivity thresholds, the smaller is the difference of POFD in the forecasts of the different setups till there is no difference at all. The differences in POD are fairly small between all setups except “Low, with filter”. POD of “Low, with filter” is lower than for “control” by up to about 0.1 at the highest reflectivity thresholds (Figure 3.9). The mentioned bias of too many “yes” events decreases with increasing reflectivity thresholds. Only the forecasts of the setup “Low, with filter" are for a reflectivity threshold of 40 bias-free, at thresholds above 40 the bias switches towards too few “yes”-events. At thresholds above 40, CSI of “Low, with filter” is lower than for the other setups, as depicted in Figure 3.10.

For short lead times, the forecasts of setup “2-layer, with filter” are performing similarly bad as the forecasts of “Low, with filter”; for greater lead times the forecasts of “2-layer, with filter” performs second worst.

Overall, the forecasts of control and "Low, no filter" are performing best with the highest POD, the lowest FAR and POFD, and the smallest bias.

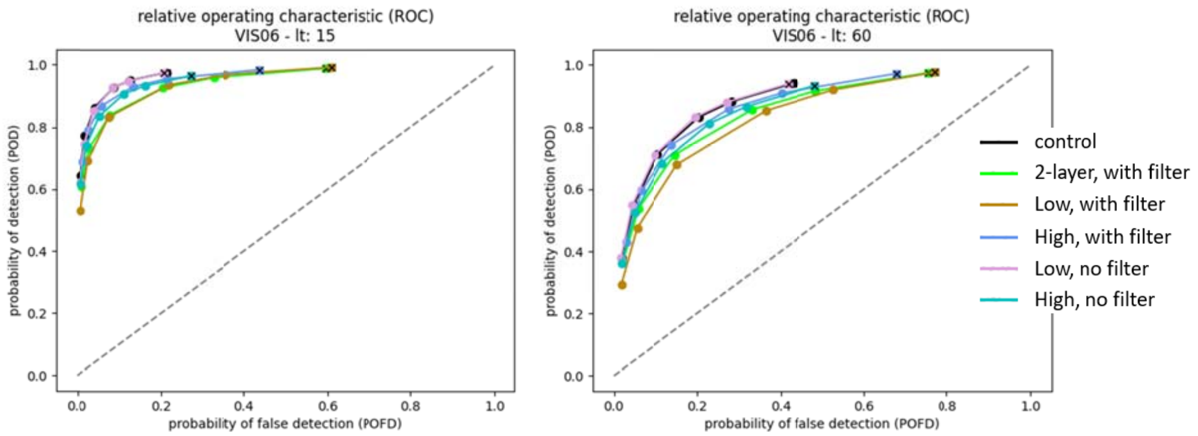


Figure 3.9 ROC curve for VIS06. Left: lead time 15, right: lead time 60. Threshold range: 9, 15, 20, 30, 40, 50. Each threshold 9 is marked with X for orientation.

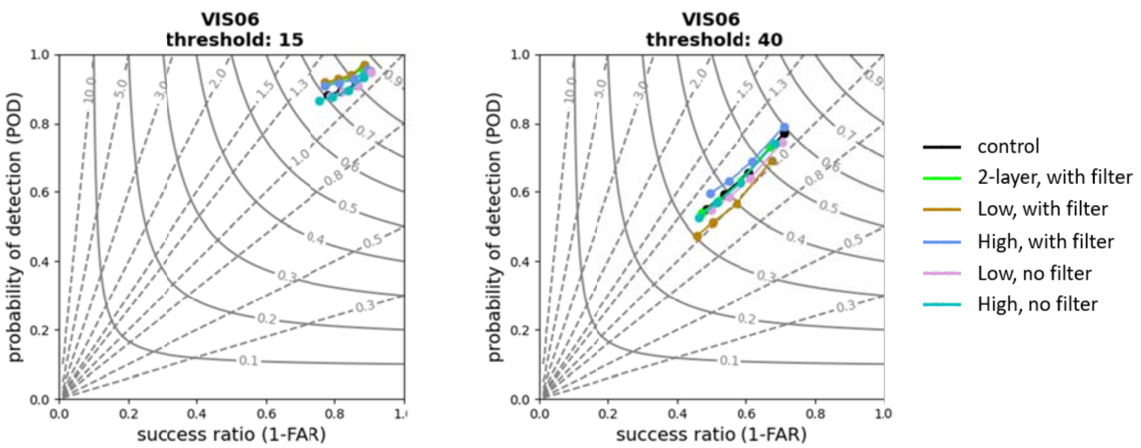


Figure 3.10 Performance diagram for VIS06. Left: for reflectivity threshold 15, right: for reflectivity threshold 40.

3.2.4 Visible Channels: VIS 08

The visible channel with wavelength *0.8 µm (VIS08)* is evaluated for the threshold range reflectivity=9, 15, 20, 30, 40, 50. Similar to *VIS06*, *VIS08* has been evaluated only at day time, defined as 06 – 18 UTC.

The forecasts of the different setups for *VIS08* show very similar behaviour in their scores as they do for *VIS06*. The control setup and “Low, no filter” are the two superior setups. At low reflectivity thresholds, the forecasts of setups with filter have an increased POFD compared to control. POD is not at all or only slightly affected. Only for forecasts of setup “Low, with filter"

POD is decreased by up to 0.1 (see Figure 3.11).

From Figure 3.12 one can tell, there is a bias towards too many “yes”-events in all setups at lead time 45 and 60.

Similarly to *VIS06*, forecasts of setup “2-layer, with filter” cannot compete with forecasts of the control run.

Forecasts of the control setup and "Low, no filter" have overall the best scores of all setups.

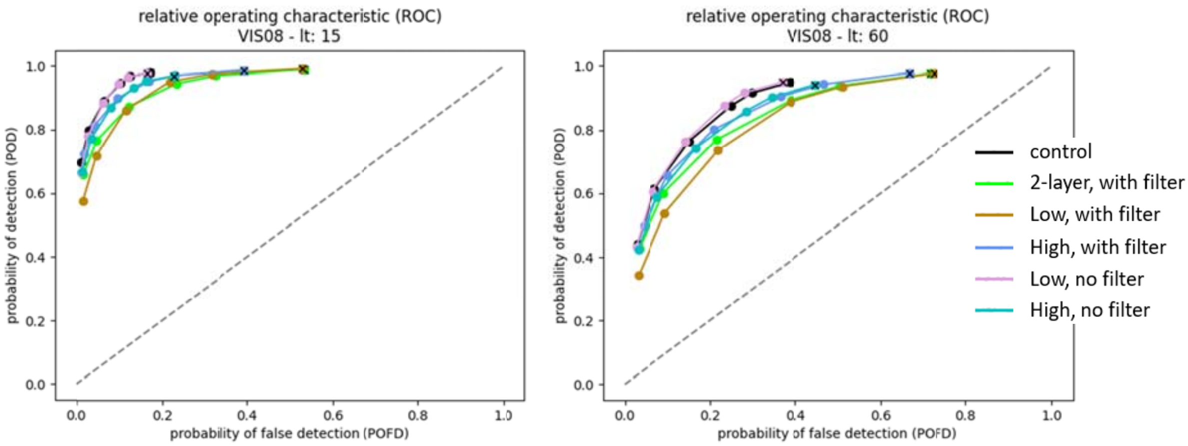


Figure 3.11 ROC curve for VIS08. Left: lead time 15, right: lead time 60. Threshold range: 9, 15, 20, 30, 40, 50. Each threshold 9 is marked with X for orientation.

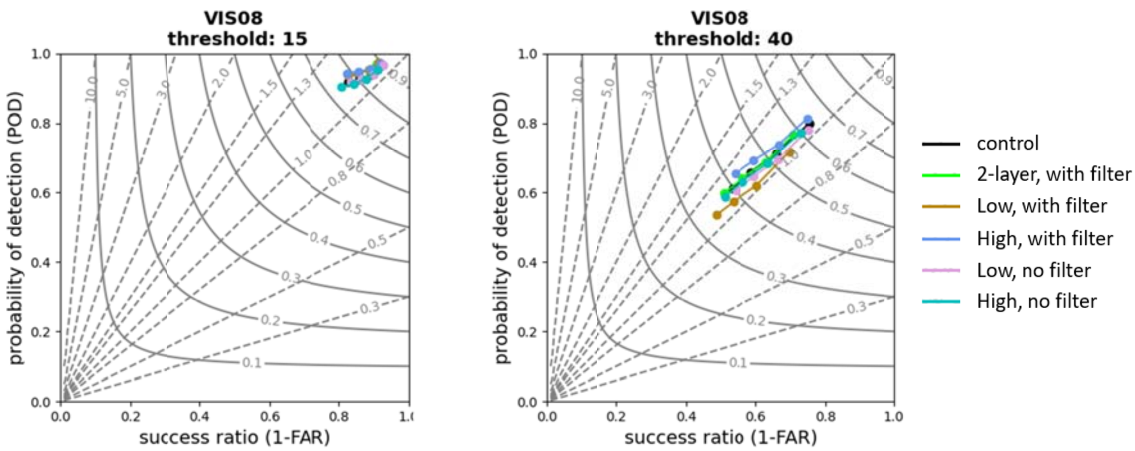


Figure 3.12 Performance diagram for VIS08. Left: for reflectivity threshold 15, right: for reflectivity threshold 40.

3.2.5 CMA: Cloud Mask

The product Cloud Mask (*CMA*) has the two categories, cloudy (1) and non-cloudy (0).

The evaluation revealed that the CTTH filter cannot be evaluated for *CMA*. As not all pixels have a height assigned, they cannot be allocated to one or another layer. Cloudy pixels do have a dedicated height and will therefore be extrapolated. However, non-cloudy pixels do not have a dedicated height. As a consequence, pixels without height assignment will be rejected in the height filtering process. This results in a complete lack of “correct negatives” (d) and “misses” (c). Without counts in the categories c and d, scores cannot be computed. Therefore, the model versions "Low, with filter" and "High, with filter" will not be listed in this chapter.

The comparison of the remaining setups shows that the forecasts of control setup achieves the best scores together with the forecasts of "Low, no filter". (see Figure 3.13). *CMA* seems to be more strongly driven by low AMVs since the setup "High, no filter" has overall slightly worse scores than control and "Low, with filter", even though the differences are marginal.

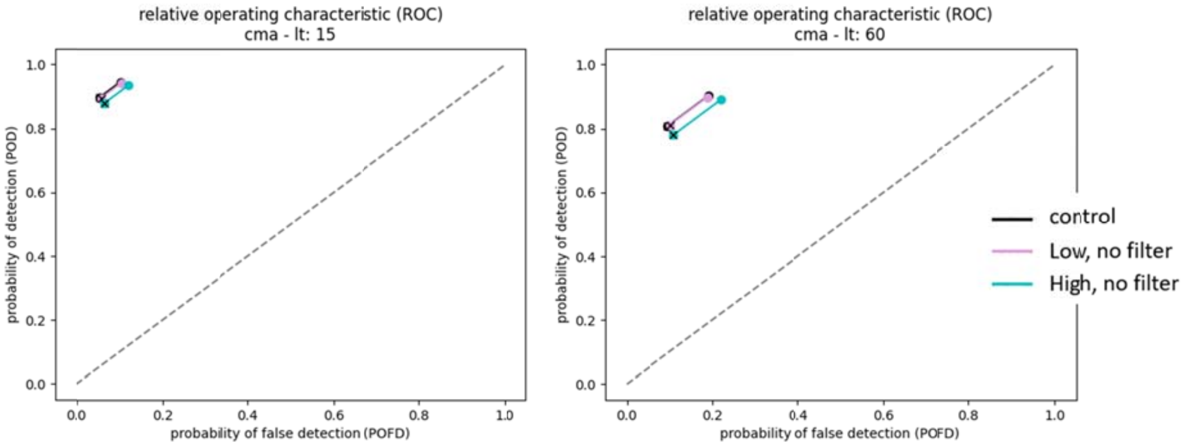
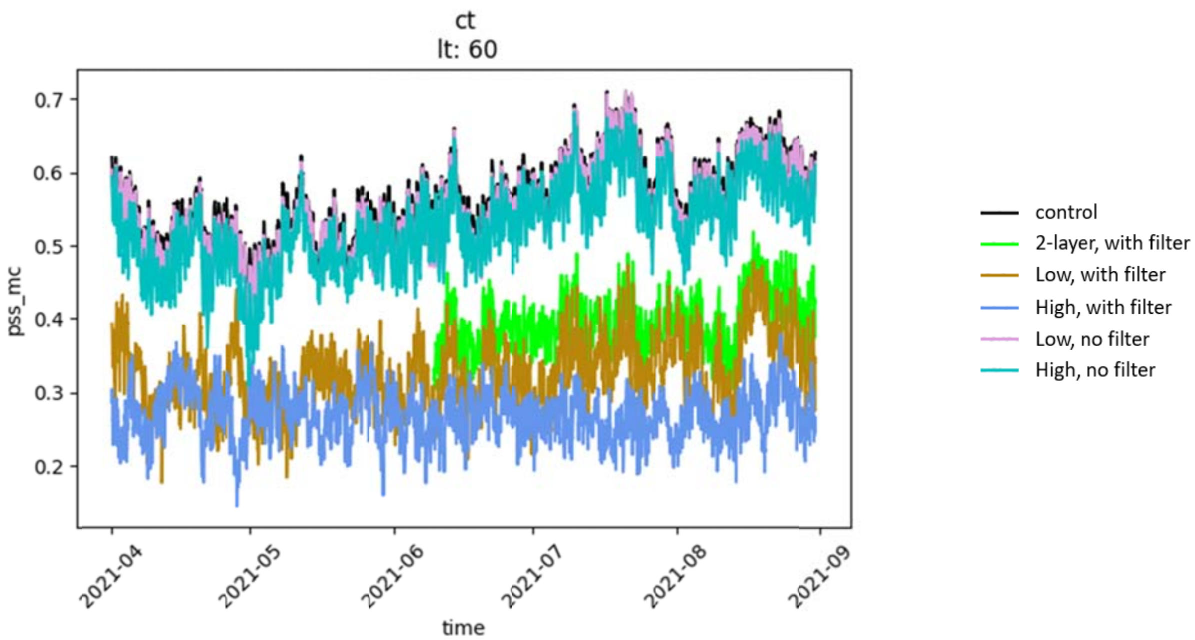


Figure 3.13 ROC curve for CMA. Left: lead time 15, right: lead time 60. Each threshold 0 is marked with X *for orientation.*

3.2.6 CT: Cloud Type

Product Cloud Type (*CT*) is one of the two multi-categorical products in this report. The categories are cloud-free land (1), cloud-free sea (2), snow over land (3), sea ice (4), very low clouds (5), low clouds (6), mid-level clouds (7), high opaque clouds (8), very high opaque clouds (9), fractional clouds (10), high semi-transparent thin clouds (11), high semi-transparent moderately thick clouds (12), high semi-transparent thick clouds (13), high semi-transparent above low or medium clouds (14), high semi-transparent above snow ice (15). The multi-categorical score PSS_{mc} has been computed in addition to the other scores.
With the holistic point of view of this evaluation, the time series of PSS_{mc} (Figure 3.14) allows the conclusion that "control" continues to be the best option for this product. Neither forecast of the other setups outperforms "control" in the multi-categorical score. A look at the scores listed by all categories separately shows a more comprehensive picture. Forecasts of "Low, with filter" improve POD (Figure 3.15) at low cloud types (5 - 7), while forecasts of all other categories are zero. And forecasts of "High, with filter" improves POD at high cloud types (e.g. 11, 12, 14), while other categories are zero or close to zero. CTs from the respectively other layer than the setups' are rejected. Furthermore, some of the CTs cannot be assigned to one of the two levels at all. The surface types (1, 2, 3 and 4) get lost in the forecast of all setups with filter, as well as category 10 which has no dedicated height. And even though singular CTs might be improving with the usage of the filter, the majority of the CTs degrade in either one of the options.

Setup "2-layer, with filter" 'cherry-picks' the improvements from the two other filter setups and forecasts of categories 5 - 7, 11, 12, and 14 are even to or gain skill compared to control (Figure 3.15). However, there are categories that not only degrade but have a complete fall-out in the forecasts of "2-layer, with filter", such as categories 1 - 4 or 10. The performance of the setups highly varies for the different cloud types. Only if a user has an interest in specific clouds types, one of the filter setups could become a reasonable choice.



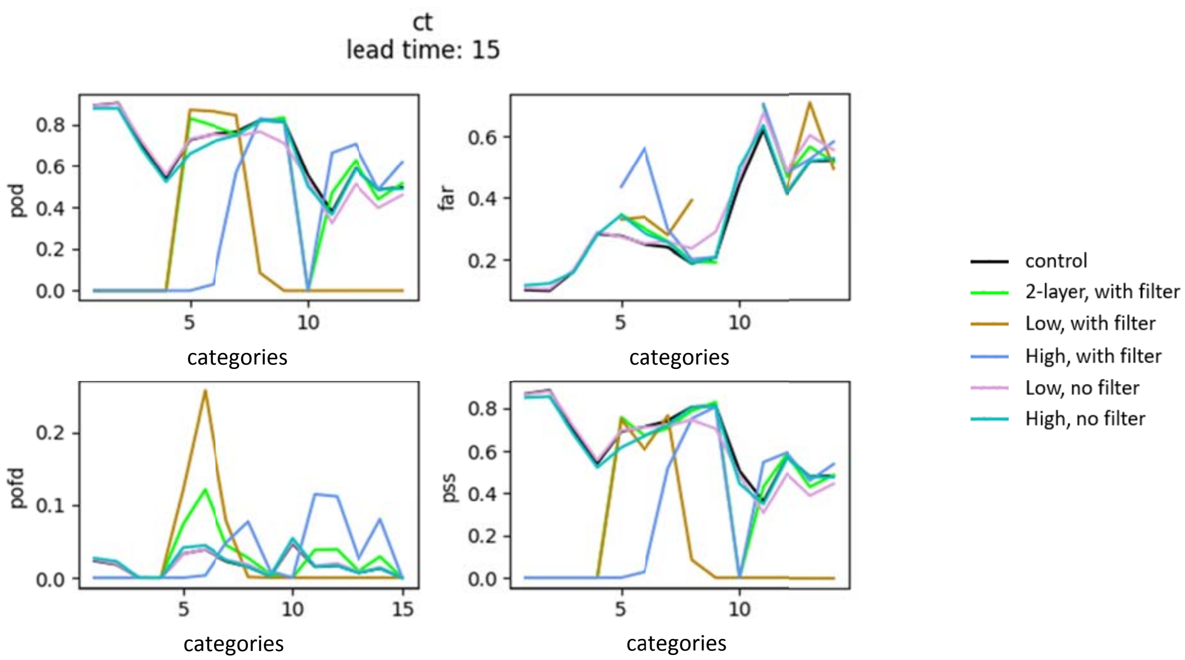


Figure 3.15 Four scores for CT listed per category at lead time 15. Top left: POD, top right: FAR, Bottom left: POFD, bottom right: PSS.

3.2.7 CTTH: Cloud Top Temperature and Height

3.2.7.1 Altitude (CTTH alti)

Product Cloud Top Temperature and Height Altitude (*CTTH alti)* is evaluated for the height range of 1000 m to 10 000 m with 1000 m intervals.

The forecasts of *EXIM* for *CTTH alti* are strongly responding to the usage of the filter, as clearly visible in Figure 3.16. For forecasts of "Low, with filter", there is only little skill in POD and PSS left above thresholds of 5000 m with both dropping to 0.5 and 0.4, respectively. No pixels at all are extrapolated by *EXIM* for setup "Low, with filter" at thresholds greater than 6000 m, POD, POFD and PSS are zero and FAR doesn't have any values. Conversely, POFD of setup "High, with filter" only starts to sink below 1 above thresholds of 5000 m and PSS starts to gain skill at thresholds greater 7000 m. This result is expected and confirms the functionality of the filtering process.

Even though the forecasts of "Low, with filter" and "High, with filter" never perform best, the forecasts of "2-layer, with filter" does compete very well with forecasts of "control". PSS and FAR of "2-layer, with filter" have a similar skill as control, which also confirms the expectations towards the filtering process.

The quality drop visible in scores of "Low, with filter" and "2-layer, with filter" at the lowest threshold comes from the chosen lower boundary of 900 hPa. Some pixels at this height get lost since they are close to the surface and below the chosen lower boundary. In contrary, the setups without filter extrapolate all pixels of literally all available heights.

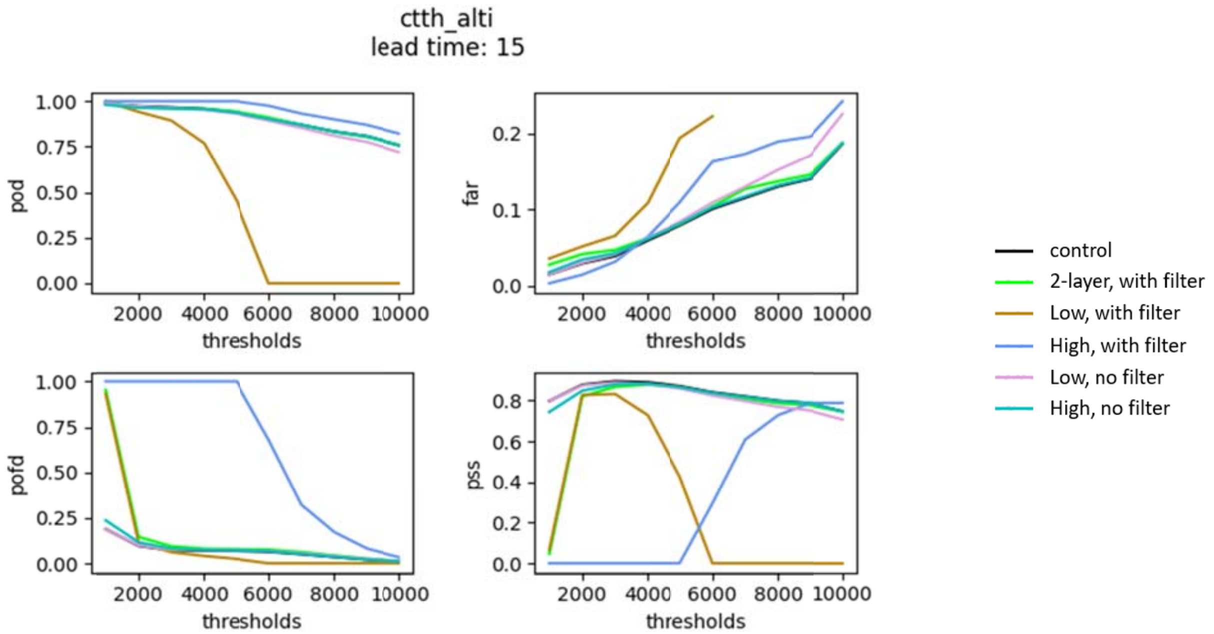


Figure 3.16 Four scores for CTTH alti listed per category at lead time 15. Top left: POD, top right: FAR, Bottom left: POFD, bottom right: PSS.

3.2.7.2 Effective cloudiness (CTTH effectiv)

Product Cloud Top Temperature and Height Effective cloudiness (*CTTH effectiv*) is evaluated in the range of 0.1 to 0.9 with a step width of 0.1.

Forecasts of *EXIM* for *CTTH effectiv* are strongly height dependent.

Skill of POFD improves for forecasts of "High, with filter" compared to control by an average of 0.2 and slightly less at thresholds greater 0.6. POD of "High, with filter" has lower skill than control and is continuously decreasing with increasing thresholds up to a maximum of 0.15. The skill of PSS is improved by about 0.15 at thresholds up to 0.5 compared to control. The improvement is getting less pronounced at thresholds above 0.5 and at thresholds above 0.8 skill of PSS is lower compared to control (Figure 3.17). Skill of FAR for “High, with filter” is slightly increased at thresholds above 0.6 with a maximum of 0.05. With a look at the performance diagram (Figure 3.18), one can see a bias of too few “yes” events for the setup "High, with filter". As POFD can artificially be improved by issuing fewer “yes” forecasts to reduce the number of false alarms, the improvement is less exciting. However, from Figure 3.18 one can conclude that the forecast bias of “High, with filter” is increasing with increasing thresholds of effective cloudiness, while skill of POFD at the very same thresholds decreases. Therefore, not all of the reduction of POFD can be invalidated by the bias.

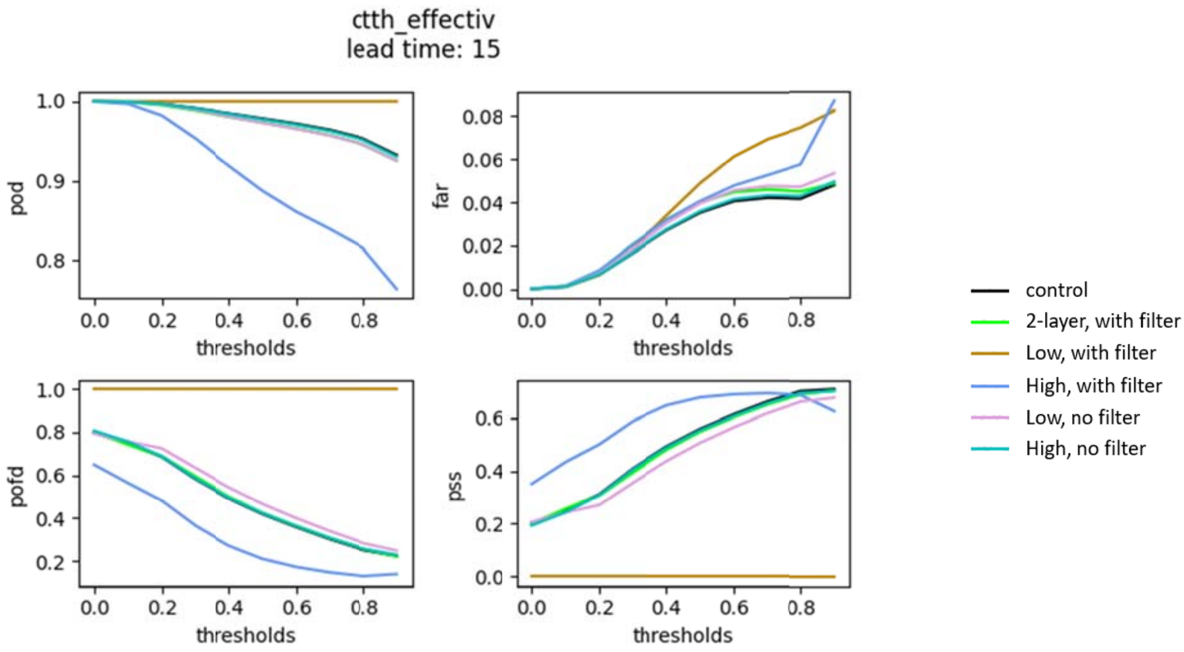


Figure 3.17 Four scores for CTTH effectiv listed per category at lead time 15. Top left: POD, top right: FAR, bottom left: POFD, bottom right: PSS. POD and POFD of setup “Low, with filter” being 1 at all thresholds implies that only values of the greatest threshold (> 0.9) are extrapolated (and therefore assigned to) the low layer.

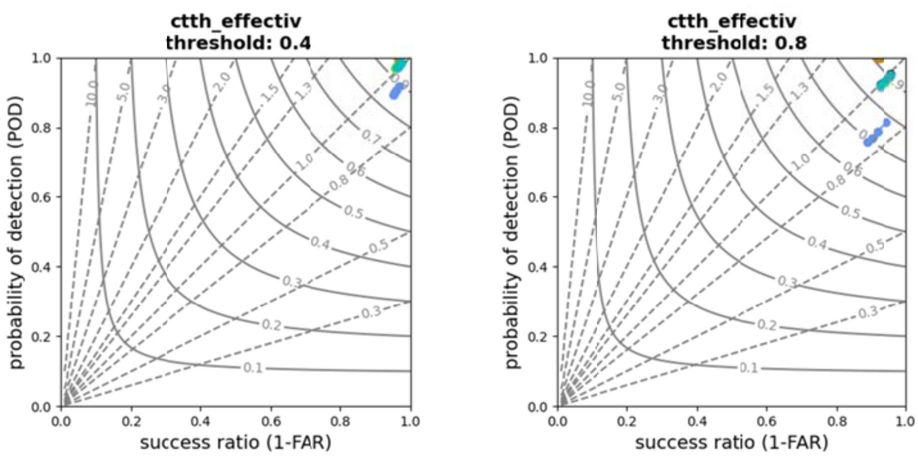


Figure 3.18 Performance diagram for CTTH effective. Left: Effective cloudiness of 0.4, right:

effective cloudiness of 0.8.

3.2.8 CMIC: Cloud Microphysics

The product cloud microphysics phase (*CMIC phase*) is the second of the two multi-categorical products. The categories are liquid (1), ice (2), mixed (3), cloud-free (4), and un-defined (5). The multi-categorical score PSS_{mc} has been computed in addition to the other scores.

The results of the three setups with filter show a degradation compared to "control" in the time series of score PSS_{mc}, as depicted in Figure 3.19. Forecasts of the two setups "Low, with filter" and "High, with filter" degrade by an order of 0.3 and 0.6, respectively, and forecasts of "2-layer, with filter" are on average lower by 0.2.

Scores listed per category are shown in Figure 3.20. POD, POFD and PSS are 0 in category 4 and 5 for all three setups with filter. Those two categories do not have assigned heights. Therefore, pixels of those categories will not be extrapolated when the filter is switched on. But also in categories 1 and 2 "control" continues to perform best. For setups "Low, with filter" and "High, with filter" skill of PSS and FAR decreases strongly compared to "control" and for setup "2-layer, with filter" skill of PSS and FAR decreases slightly. In category 3, skill of "2-layer, with filter" is similar to "control". Skill of FAR and PSS is worse compare to control for "Low, with filter" and "High, with filter", respectively.

With the holistic point of view of including all categories and not selecting specific categories, "control" is the superior option.

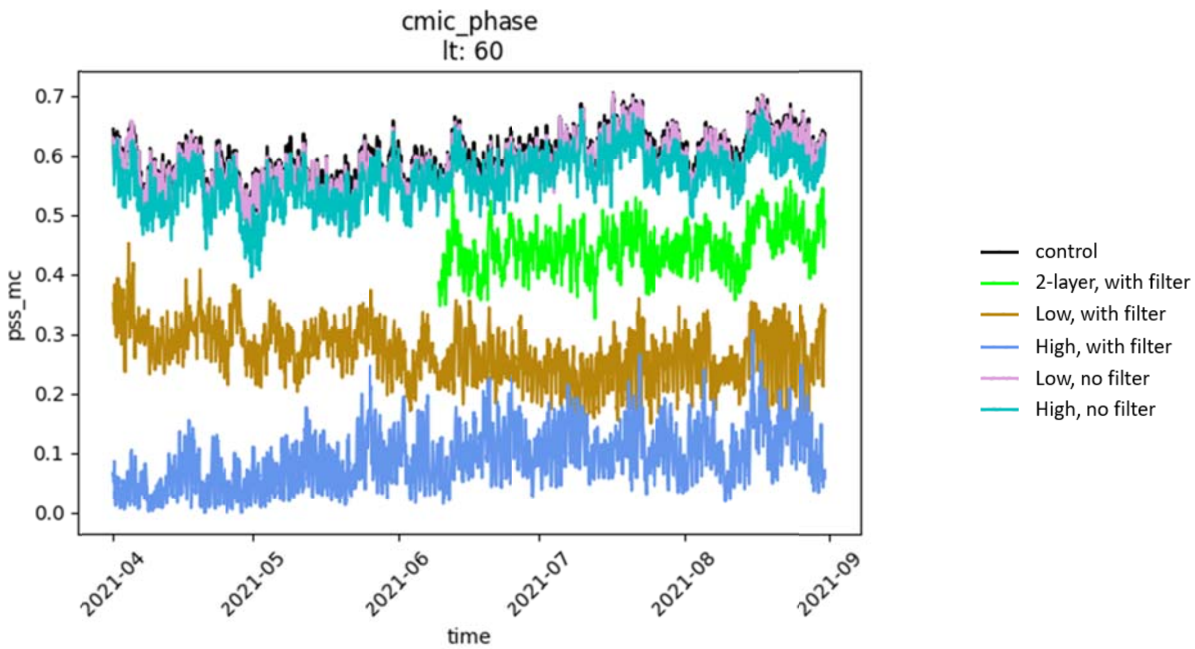


Figure 3.19 Time series of PSS_{mc} for CMIC phase at lead time 60.

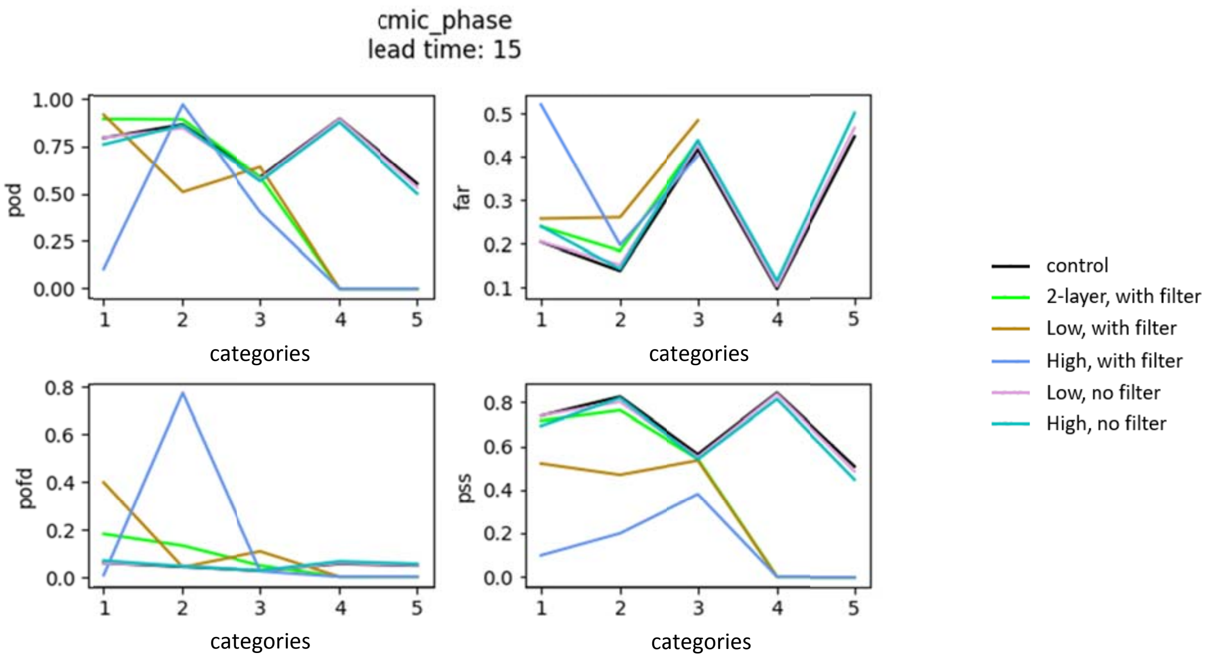


Figure 3.20 Four scores for CMIC phase listed per category: liquid (1), ice (2), mixed (3), cloud-free (4), and un-defined (5) at lead time 15. Top left: POD, top right: FAR, Bottom left: POFD, bottom right: PSS.

3.2.9 CRR: Convective Rainfall Rate

The *EXIM* forecasts of Convective Rainfall Rate (*CRR*) are analysed in the range of 1 to 5 mm/h with an interval of 1 mm/h.

CRR is driven by high-level events. This is reflected in the results depicted in Figure 3.21 and Figure 3.22.

Forecasts of “High, with filter”, “High, no filter” and “control“ have fairly the same scores, in particular POD, FAR and POFD don’t differ between the three setups. Only POFD of setup “High, with filter” is marginally increased compared to control at lead time 60 and for small thresholds. This increase of POFD might be caused by a reduced amount of total pixels. On the other hand, scores of the setup “Low, with filter”, which ignores high layer information, hugely degrade (Figure 3.21 and Figure 3.22). Skill of POD decreases by up to 0.4, FAR increases by up to 0.3 and CSI reduces by more than 0.3. There is a huge bias for setup “High, with filter” towards too few “yes”-events. Forecasts of the setup “Low, no filter” have very little skill. With high-layer pixels also being extrapolated, the scores don’t degrade as much as for “Low, with filter”. Still, there is a loss of skill in POD, FAR and CSI by about 0.1 (Figure 3.22).

Forecasts of setup “2-layer, with filter” have very similar scores as the control setup. POD and CSI are marginally higher for “2-layer, with filter” than for “control” which hints that there are few low-layer precipitation pixels being correctly extrapolated by the low layer.

All scores degrade with increasing CRR intensity due to fewer occurrences.

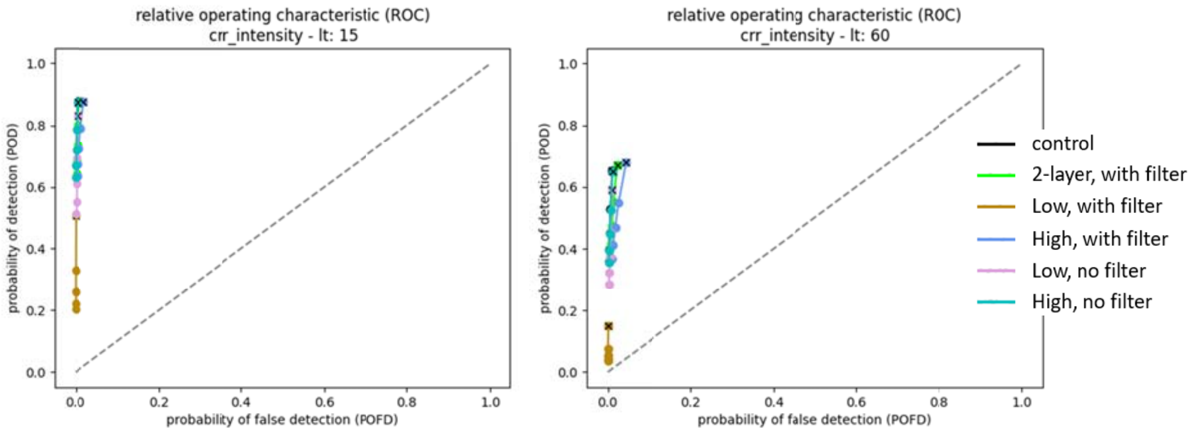


Figure 3.21 ROC curve for CRR. Left: lead time 15, right: lead time 60. Threshold range: 1, 2, 3, 4, 5 mm/h. Each threshold 1 is marked with X for orientation.

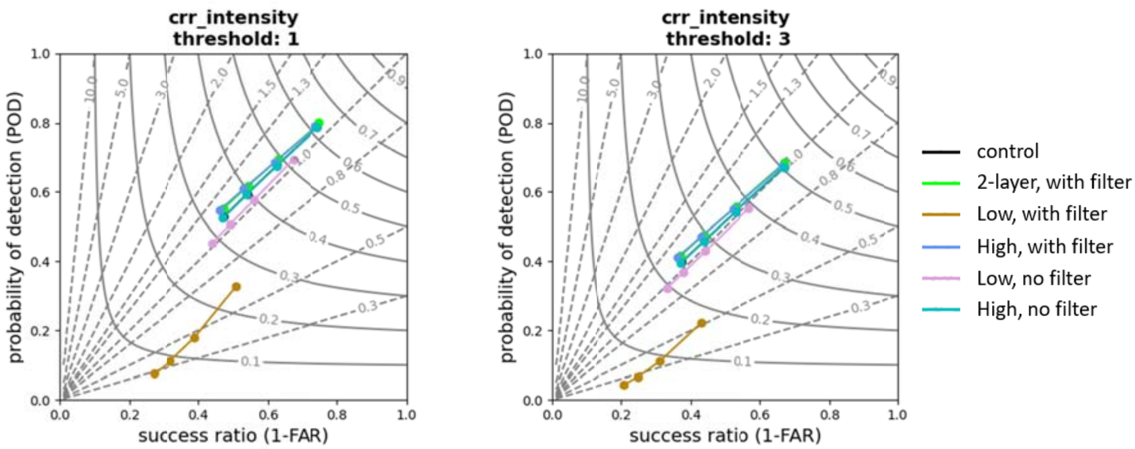


Figure 3.22 Performance diagram for CRR. Left: precipitation intensity 1 mm/h, right: precipitation intensity 3 mm/h.

3.2.10 CRRPh: Convective Rainfall Rate from Cloud Physical properties

EXIM forecasts of Convective Rainfall Rate from Cloud Physical properties *(CCR-Ph)* are evaluated for the thresholds 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, 21.0 mm/h.

The forecasts of the various setups for *CRRPh* show a similar behaviour as they do for *CRR*. The product can be associated with the high level, as illustrated in Figure 3.24. Forecasts of “High, no filter” and “High, with filter” are performing very similar as “control”. Only at lead time 60, POFD and POD of “High, with filter” are slightly increased by about 0.05 (Figure 3.23).

Without any high-level information as is the case for setup “Low, with filter”, POD, FAR and CSI are hugely degraded compared to “control” by up to 0.3, 0.2 and 0.2, respectively. In forecasts of “Low, with filter” there is also a bias towards too few “yes” events, only POFD is not affected. Also, forecasts of “Low, no filter” are degraded compared to control with POD being lower by 0.1 and CSI and FAR being worse by about 0.05 than control.

“2-layer, with filter” are relatively competitive with forecasts of “control” when predicting lower precipitation intensities however, there is a small drop in skill of the score CSI with increasing intensity thresholds.

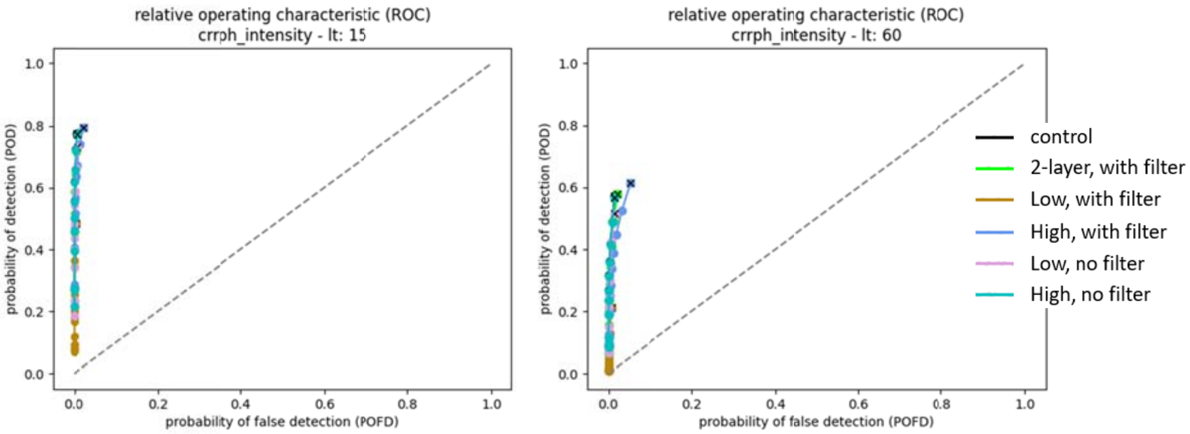


Figure 3.23 ROC curve for CRRPh. Left: lead time 15, right: lead time 60. Threshold range: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 21 mm/h. Each threshold 1 is marked with X *for orientation.*

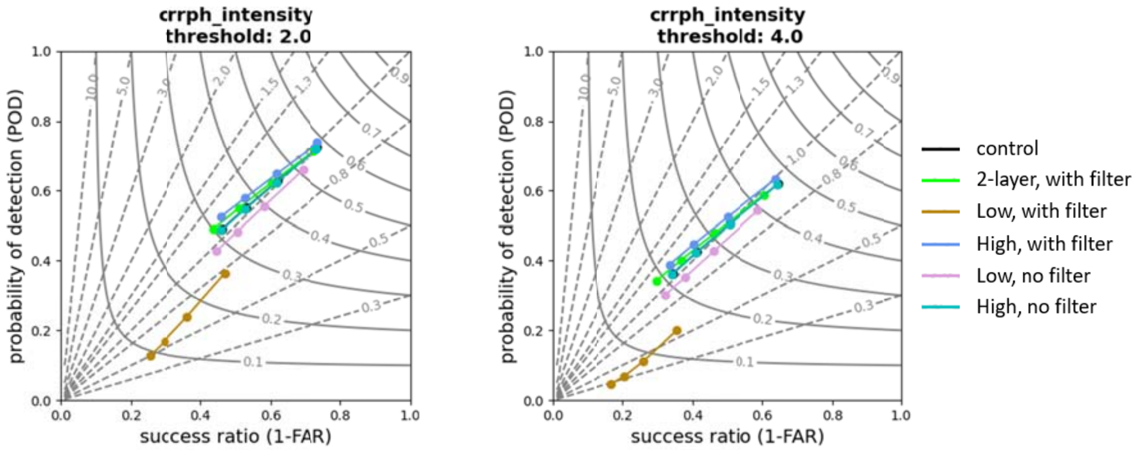


Figure 3.24 Performance diagram for CRRPh. Left: Precipitation intensity 2 mm/h, right precipitation intensity 4 mm/h.

3.2.11 PC: Precipitating Clouds

The *EXIM* forecasts for the product Precipitating Clouds (*PC*) is evaluated for 5 precipitation likelihoods: 10 % (1) 20 % (2), 30 % (3), 40 % (4), 50 % (5). *PC* does not detect precipitation from low clouds.

The setup “Low, with filter" (Figure 3.26) loses skill compared to “control” for the scores POD, FAR and CSI. POD degrades compared to “control” for the lowest threshold by about 0.1. For higher thresholds the degradation is pronounced more strongly. POD degrades compared to “control” by 0.2 and up to 0.3 with increasing lead time. Skill of CSI is reduced by about 0.1 at threshold 1. The degradation deepens for greater thresholds to up to 0.2 at threshold 5.

Skill of POD, FAR and CSI for “2-layer, with filter” is fairly the same at threshold 1 as the control setup’s skill. However, with increasing thresholds, POD, FAR and CSI are degrading by up to 0.1 (Figure 3.26).

For setup “High, with filter”, values of POFD are increasing at the lowest two thresholds (Figure 3.25). CSI of “High, with filter” is the highest CSI of all setups at threshold 1 for lead

times greater 15, the difference is getting more pronounced with increasing lead times. For greater thresholds the improvement of CSI does not apply (Figure 3.26).
With increasing precipitation likelihoods, FAR and CSI of "High, with filter" are slightly worse than "control" or the no-filter options but with a more neutral bias.

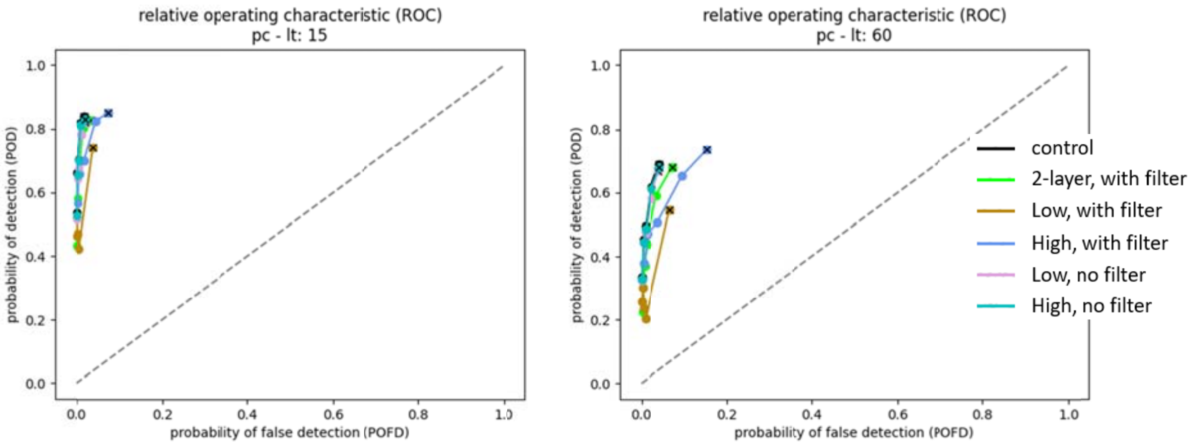


Figure 3.25 ROC curve for PC. Left: lead time 15, right: lead time 60. Threshold range: 1, 2, 3, 4, 5. Each threshold 1 is marked with X *for orientation.*

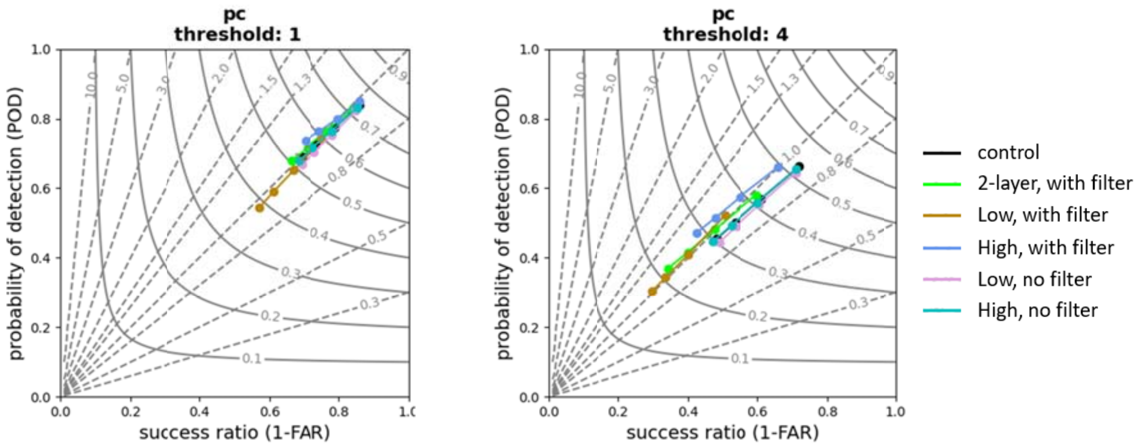


Figure 3.26 Performance diagram for PC. Left: for threshold "precipitation likelihood category 1" (10 %), right: for threshold "precipitation likelihood category 4" (40 %).

3.2.12 PCPh: Precipitating Clouds from Cloud Physical Properties

The *EXIM* forecasts of the product Precipitating Clouds from Cloud Physical Properties (*PCPh*) are evaluated for precipitation probabilities 11, 21, 31, 41, 51, and 61 %. Due to the nature of the product, only daytime has been considered for this analysis. Daytime is defined as 06 – 18 UTC.
The scores of the setups for *PCPh* show similar behaviour as for *CRRPh* and are dominated by high-level weather. This can be seen in Figure 3.28 where the skill drops most pronouncedly for the setup without high-level pixels. "Low, with filter" performs way worse than all of the other setups. POD of "Low, with filter" decreases compared to "control" by about 0.2 at lead time 15. With increasing lead time, the degradation deepens further and POD decreases by up to 0.3. Skill of FAR increases by the same order of magnitude as the skill of POD degrades (from 0.2 to a max of 0.3). With increasing precipitation probabilities, forecasts of setup "Low, with filter" develop a bias towards too few "yes"-events which further deepens with rising precipitation probabilities. The bias also deepens with increasing lead time.

The forecasts of “Low, no filter” have slightly lower values of POD than “control” by about 0.05. Forecasts of “High, with filter” result in higher POFD (Figure 3.27) in the low probability range by up to 0.05 at lead time 60. This is most likely due to a smaller number of correct negatives, lacking pixels from the low-level. POD and CSI of “High, with filter” are also increased by about 0.05 at lead time 60.
Scores of “High, no filter” and “2-layer, with filter” don’t differ as much from “control” as the three versions described before. Depending on the specific probability threshold and lead time, one time “High, with filter” scores best, another time the “2-layer, with filter” together with “High, no filter".

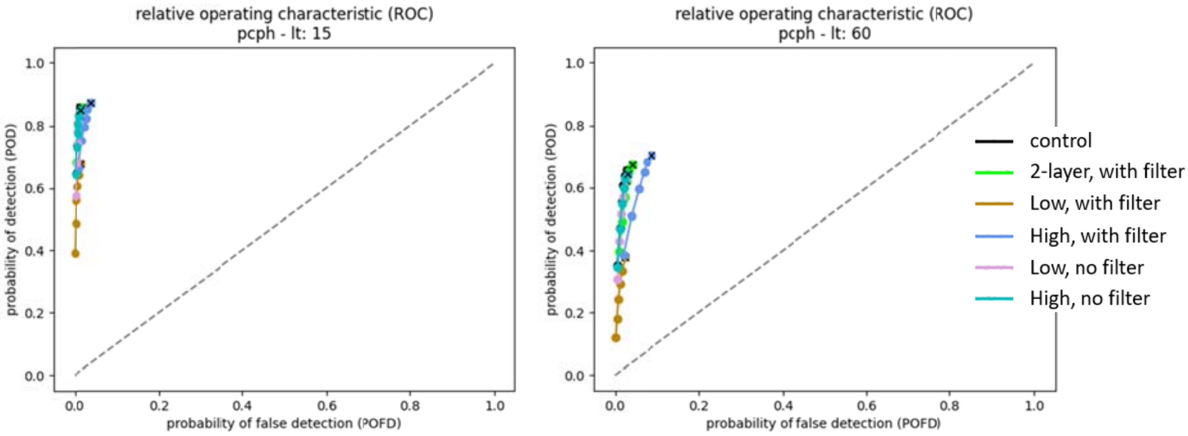


Figure 3.27 ROC curve for PCPh. Left: lead time 15, right: lead time 60. Threshold range: 11, 21, 31, 41, 51, 61 %. Each threshold 11 is marked with X *for orientation.*

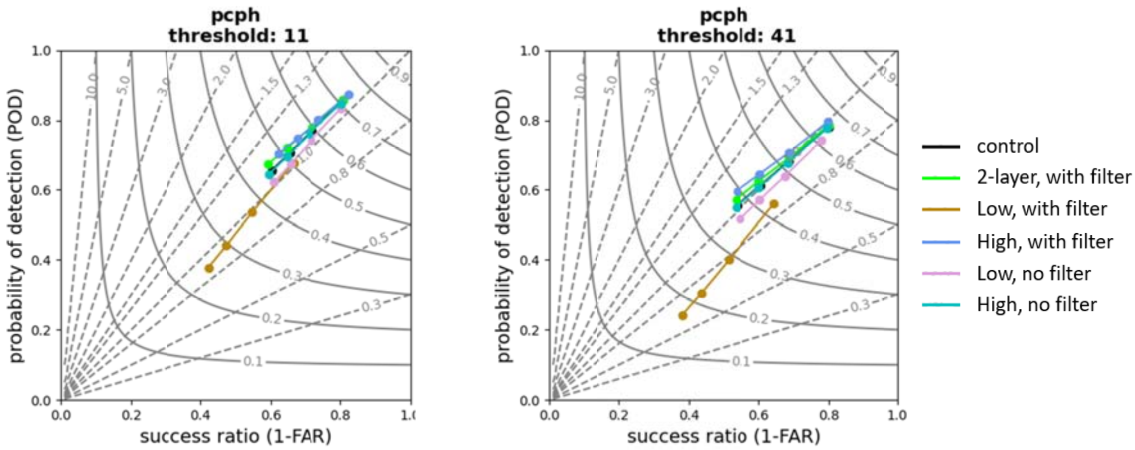


Figure 3.28 Performance diagram for PCPh. Left: for threshold precipitation likelihood=11 %, right: for threshold precipitation likelihood=41 %.

3.3 EVALUATION OF INCLUDING WV VECTORS

This chapter discusses the results of the evaluation whether water vapour (WV) atmospheric motion vectors (AMVs) should be included to the current set of used AMVs for extrapolation of all products and not just for the WV channels.

The results show that the products fall into two groups. In group one there are some thresholds and/or lead times where “setup with WV AMVs” achieves better results than “control, without WV AMVs”. In group two the “control, without WV AMVs” achieves better results in the majority of the cases for all thresholds and lead times. But most notably, there is no product where "setup with WV AMVs" generally improves the scores for all thresholds and lead times.

The qualitative changes caused by WV AMVs are small and for the majority of the products negative.

The relatively greatest improvements are gained for the products *CRR*, *CRRPh* and *IR108* (mentioned as group one above). Taking *CRR* as an example for products with one of the most improving scenarios, "setup with WV AMVs" results in the majority of the cases having better POD scores than "control, without WV AMVs". See in Figure 3.29 that scores improve in more than 70 % of all cases for intensity threshold 1 and at all lead times. At lead time 60 the amount of improving cases is lower with 65 %. For threshold 2 and 3 and at lead time 15, 85 % of the cases improve. With increasing lead time, the amount of improved cases gets smaller. Skill of PSS improves with a similar pattern as POD but with slightly lower percentage values, ranging from 60 % to a max of 80 %. POFD and FAR are mainly get worse. Solely for the highest thresholds at short lead times, the majority of cases get better instead of worse for the scores POFD and FAR. With decreasing intensity thresholds and increasing lead time "control, without WV AMVs" continues being the better choice in the majority of cases. Quantitatively, the changes are very small. The biggest improvements for POD and PSS are of the order of 0.008 and the biggest degradations for FAR are 0.006.

CMA, *VIS06* and *VIS08* belong to group two. Their scores are degrading when including WV AMVs. Taking *VIS08* as an example, one can see in Figure 3.30 that the "control, without WV AMVs" is better compared to "setup with WV AMVs" in the majority of the cases. The scenarios for which the scores of the "setup with WV AMVs" are best don't even reach a third of improving cases. In the worst scenarios, only about 0.05 % of all cases improve. The qualitative changes are small, for POD the difference between the setups are up to a degradation of 0.006 and for FAR about 0.008. Skill of PSS degrades most strongly with maximum values greater 0.01.

There is no clear improvement for any of the products and only some show minor enhancements for some categories and lead times.

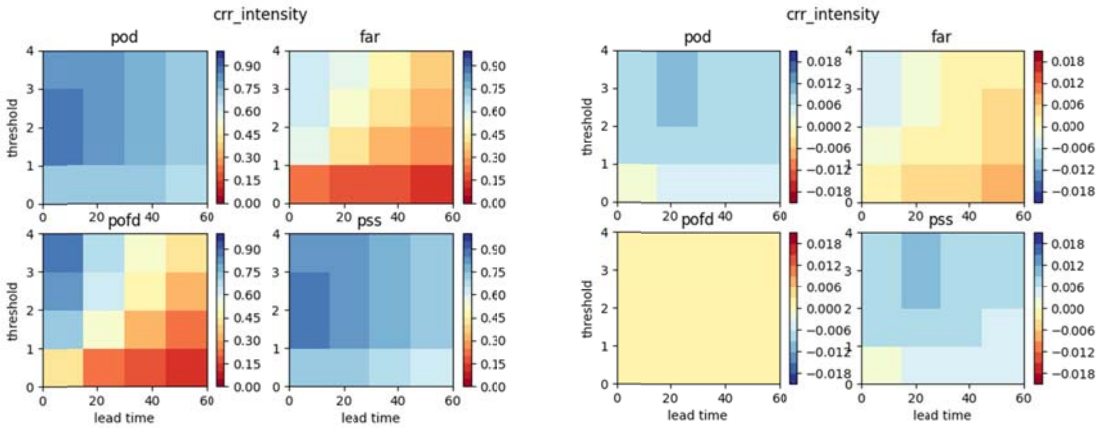


Figure 3.29 Left: Frequency of cases where including WV AMVs for CRR yields scores greater than its control. Right: Amount of change in scores, after including WV AMVs for forecasting CRR. Both figures: Lead time vs. threshold categories. Blue values show an improvement, yellow improvements and degradations of the same amount, red a degradation.

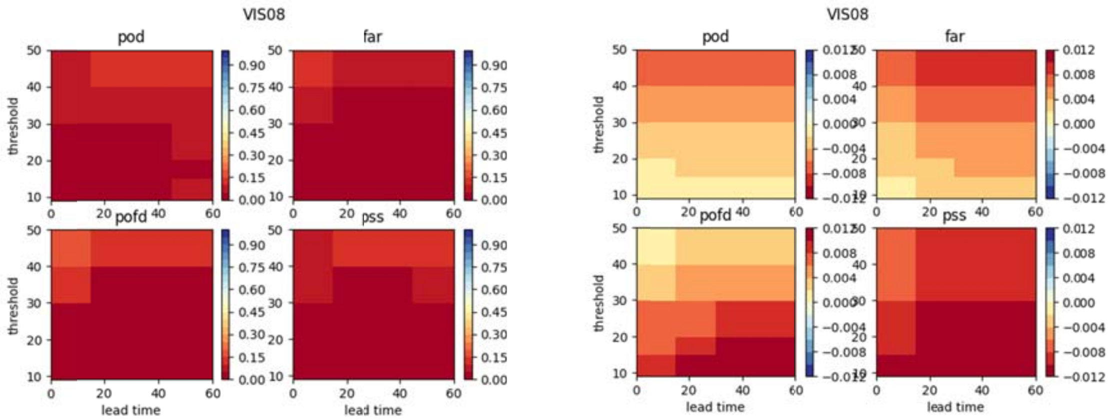


Figure 3.30 Left: Frequency of cases where including WV AMVs for VIS08 yields scores greater than its control. Right: Amount of change in scores, after including WV AMVs for forecasting VIS08. Both figures: Lead time vs. threshold categories. Yellow shows improvements and degradations of the same amount, red a degradation.

4. CONCLUSIONS

This validation report discusses three main aspects. First, whether the two modified products and the newly added can beneficially be extrapolated by *EXIM*. Second, how the newly introduced CTTH filter performs compared to previous extrapolation setups. And third, whether water vapour atmospheric motion vectors shall be added to the list of used AMVs. Summary conclusions based on the detailed results shown in Chapter 3 are presented hereafter.

4.1 EXIM OUTPERFORMS PERSISTENCE

The forecasts of *EXIM* are validated with each change in the algorithm of an input product or when a new product is added to the set of extrapolated products in *EXIM*.

The underlying results confirm that forecasts of *EXIM* add value for all three investigated products (*CTTH effectiv*, *PCPh*, *CRRPh*) compared to persistence, and they all reach the threshold accuracy of being "*on average better than persistence forecast*".

POD of *CTTH effectiv* improves skill compared to persistence for transparencies greater 0.1. POFD and FAR at thresholds greater 0.3 (excluding lead time 60 where the minimum threshold for improvements is 0.4). Therefore, the forecasts of *EXIM* for *CTTH effectiv* are overall better than persistence for an effective cloudiness greater 0.3.

The forecasts of *EXIM* for *PCPh* improve POFD and FAR from precipitation probabilities greater 11 %. Forecasts of *EXIM* for the product *PCPh* are on average better than persistence forecast except for very low precipitation likelihoods.

While the forecast of *EXIM* for *CRRPh* improves FAR and POFD for all precipitation intensities compared to persistence, POD only improves for intensities up to 9 mm/h.

4.2 CTTH FILTER AND ITS RECOMMENDATION

EXIM has been extended by a new feature, namely the CTTH filter. The CTTH filter ensures that pixels are extrapolated only with AMVs stemming from the same defined layer. Concluding on the results from Chapter 3, the products can at first be divided into products for which the CTTH filter can be applied and products for which the CTTH filter cannot be evaluated and applied in their full extent.

CMA, *CT* and *CMIC* belong to the set of products for which the CTTH filter *cannot* be evaluated. They cannot be extrapolated in their full extent when using the CTTH filter because all three products have some categories that do not have an assigned height. Pixels of those categories will get lost by the height selection process and will therefore not be extrapolated. With pixels from some categories missing completely, the setups using the filter cannot compete with the control setup even though other categories with correct height assignments might improve.

For all other products and channels (*IR108, IR38, VIS06, VIS08, CTTH alti, CTTH effectiv, CRR, CRRPh, PC, and PCPh*) the CTTH filter *can* be applied and evaluated.

The products can be grouped as follows: *First*, "High, with filter" and "2-layer, with filter" can compete with "control", *second*, only "High, with filter" can compete with "control", *third*, only "2-layer, with filter" can compete with "control" and *fourth*, none of the setups with filter can compete with "control" for all scores.

To start with an overall observation, forecasts of setup "Low, with filter" never outperform the other setups for any of the products. Too much information gets lost when focussing on the herein chosen low-layer and with the comprehensive evaluation approach of this report.

CRR, *PCPh* and *CTTH effectiv* belong to the ***first group***. Due to the nature of the product *CRR* can be associated to the high level. The forecasts of both, "High, with filter" and "2-layer, with filter", perform very similar as the forecasts of control. None of the setups is best for all scores, lead times and thresholds and it slightly varies which setup dominates when. Only the forecasts of the two low-layer versions have clearly lower scores than "control". The scores of *PCPh* picture similar results, with the high layer versions and "2-layer, with filter" having very similar scores as "control". *CTTH effectiv* differs slightly compared to the other two products. POFD of "High, with filter" is improved over the whole threshold range, while POD decreases with increasing thresholds, resulting in an improved PSS for most of the thresholds. Forecasts of "2-layer, with filter" are as good as the ones from the control setup.

CRRPh and *PC* are allocated to the ***second group.*** *CRRPh* and *PC* are both associated with high-level events. For both products, forecasts of the two low-level setups have lower skills than the other setups. In the range of low thresholds "High, with level" and "2-layer, with layer" compete regarding their skill against "control". With increasing thresholds, "2-layer, with filter" slightly loses skill compared to "control".

IR108, *IR38* and *CTTH alti* are in the ***third group***. All three products have a strong height dependency which is reflected in the scores. Scores of the setups "Low, with filter" and "High, with filter" are highest in those altitudes their layer is associated with and lowest in other altitudes. Scores of "2-layer, with filter" are competitive with "control" in all altitudes and benefit from the improvements in both layers. When lowering the lower boundary of the lower layer, the results might benefit even more. Currently, the very low pixels get lost in the setups with filter.

VIS06 and *VIS08* are in the ***fourth group.*** In none of the setups with filter all scores of the forecasts improve compared to "control".

In general, for all products there is one practical aspect that confers setups without filter an advantage. In cloud-free regions, there is no CTTH available and therefore there is no height data. Pixels will be rejected. For no filter setups there is a post-processing step interpolating those data gaps. On the other hand, for setups with filter, the post-processing step is switched off in order to keep the clean separation between the two layers. On average, the strict rule "no pixel without height assignment from analyses" causes a loss of 2/3 to ½ of all pixels in the *EXIM* products.

As a remark, the results and conclusion of this report are addressed by a holistic point of view. With a different or more specific focus, there might be advantages of certain configurations that have not been considered in this report. Users can choose whatever one or two layers are relevant for them and gain information not mentioned in this report.

4.3 WV VECTORS USAGE

In order to answer the question whether WV AMVs should be added to augment the vector field used for extrapolation, this evaluation has been conducted. To summarise the results from Chapter 3, including WV AMVs doesn't improve the forecasts. Even though the number of vectors increases with WV vectors, there is no clear improvement for any of the evaluated scores. Most of the product extrapolations are even getting worse. The strongest degradation can be found for *CMA*, *VIS06* and *VIS08*. Some improvements are visible for *CRR*, *CRRPh* and *IR108*, but also for those products a recommendation of using WV AMVs cannot be given, since degradations of other scores are just as big as the improvements achieved. The order of magnitude is either way rather small.

Therefore, it can be concluded that the extrapolation by *EXIM* is not improving with the usage of WV AMVs as the product movements seem to be different than the WV AMVs. And considering the additional computation capacities needed, an inclusion of the WV AMVs is not recommended.

5. OUTLOOK

This underlying validation displayed that forecasts of *EXIM* are outperforming persistence for the implemented products and satellite channels for all lead times. However, the dominance is decreasing with increasing lead time. For all moving pixels, which should be the majority, *EXIM* will get better with time. Only stationary events and at the boundaries of the domain where the movement origins from, persistence will dominate with increasing lead time. The reason why the improvement by forecasts of *EXIM* is getting smaller with increasing lead time is not fully understood. The investigation of this question is foreseen to be addressed in future validations.

6. ANNEX

6.1 ROC CURVE

The relative operating characteristic (ROC) curve (Figure 6.1) depicts the relation of POD and POFD. Therefore, PSS can be inferred, with PSS being 1 in the top left corner and decreasing vertically to the dashed line. At the dashed line PSS = 0 and the forecast has no skill.

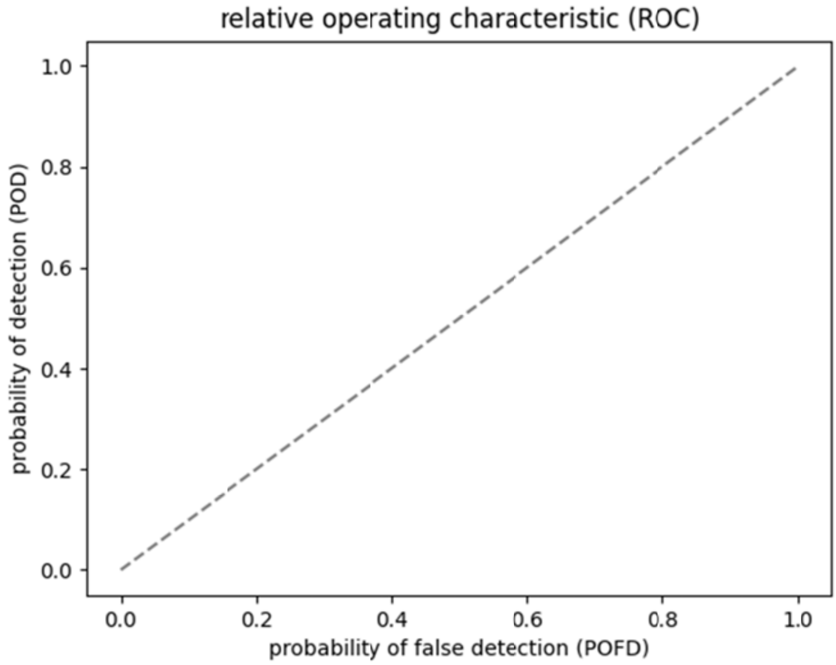


Figure 6.1 Exemplary ROC curve without data.

6.2 PERFORMANCE DIAGRAM

The performance diagram (Figure 6.2) summarises the scores POD, FAR, CSI and BIAS in one plot. A perfect forecast would be in the top right corner; the worst scenario would be in the bottom left corner.

POD		Probability of detection. Increasing from bottom to top.
FAR		False Alarm Ratio. Increasing from right to left.
BIAS	- - -	Bias. The diagonal is bias-free (1.0). Lower values express too few (forecasted) “yes”-events and higher values too many (forecasted) “yes”-events.
CSI	___	Critical Success Index. Curved lines. Increasing from bottom left corner to top right corner.

For a more comprehensive description, please refer to:
https://www.cawcr.gov.au/projects/verification/Roebber/PerformanceDiagram.html

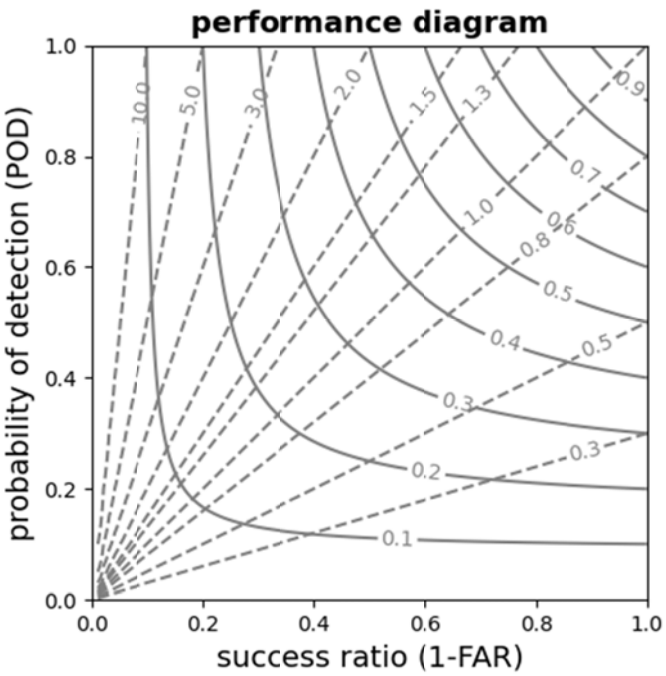


Figure 6.2 Exemplary performance diagram without data.