



# Automatic quality control of the Meteosat First Generation measurements

Freek Liefhebber<sup>1</sup>, Sarah Lammens<sup>1</sup>, Paul W. G. Brussee<sup>1</sup>, André Bos<sup>1</sup>, Viju O. John<sup>2</sup>, Frank Rüttrich<sup>2</sup>, Jacobus Onderwaater<sup>2</sup>, Michael G. Grant<sup>2</sup>, and Jörg Schulz<sup>2</sup>

<sup>1</sup>Science and Technology B.V., Delft, the Netherlands

<sup>2</sup>EUMETSAT, Darmstadt, Germany

**Correspondence:** Freek Liefhebber (liefhebber@stcorp.nl)

Received: 19 June 2019 – Discussion started: 29 July 2019

Revised: 19 December 2019 – Accepted: 6 January 2020 – Published: 10 March 2020

**Abstract.** Now that the Earth has been monitored by satellites for more than 40 years, Earth observation images can be used to study how the Earth system behaves over extended periods. Such long-term studies require the combination of data from multiple instruments, with the earliest datasets being of particular importance in establishing a baseline for trend analysis. As the quality of these earlier datasets is often lower, careful quality control is essential, but the sheer size of these image sets makes an inspection by hand impracticable. Therefore, one needs to resort to automatic methods to inspect these Earth observation images for anomalies. In this paper, we describe the design of a system that performs an automatic anomaly analysis on Earth observation images, in particular the Meteosat First Generation measurements. The design of this system is based on a preliminary analysis of the typical anomalies that can be found in the dataset. This preliminary analysis was conducted by hand on a representative subset and resulted in a finite list of anomalies that needed to be detected in the whole dataset. The automated anomaly detection system employs a dedicated detection algorithm for each of these anomalies. The result is a system with a high probability of detection and low false alarm rate. Furthermore, most of these algorithms are able to pinpoint the anomalies to the specific pixels affected in the image, allowing the maximum use of the data available.

## 1 Introduction

Earth observation (EO) from geostationary satellites provides a wealth of information which can be used to study the Earth's climate system as described, for example, by Rossow and Schiffer (1999). While their focus of interest was long-term cloud effects, other studies also have used those data for deriving information on land surface temperatures (Duguay-Tetzlaff et al., 2015), upper tropospheric humidity (Soden and Bretherton, 1993), and solar surface irradiance (Mueller et al., 2015). EUMETSAT's Meteosat First Generation (MFG) satellites were equipped with the Meteosat Visible Infra-Red Imager (MVIRI) instrument. The MVIRIs had been measuring in three distinct wavelengths: (i) visible (VIS) – providing information on surface and atmospheric albedo (Rüttrich et al., 2019); (ii) water vapour (WV) providing information on upper tropospheric humidity, which is a key climate variable, yet not very well simulated by the current climate models (John and Soden, 2007) and (iii) infrared (IR) window providing information on the surface and cloud top temperature and on the presence of clouds (Stoeckli et al., 2019; John et al., 2019). Data from Meteosat-1, launched in 1977, are available for only a year from December 1978 to November 1979. There are continuous data available from Meteosat-2 onwards starting in February 1982 until April 2017, when the last satellite in the series, Meteosat-7, was moved to its graveyard orbit.

If such an image dataset is to be used for analysing how the climate varies over time, one must control the quality of the dataset in order to avoid any bias in the result. The potential value of such a historical time series, however, is partially hindered due to the presence of image anomalies (Koeppen,

2004) (acquisition errors, physical effects such as stray light, etc.) and radiometric anomalies (Brogniez et al., 2006) (calibration errors, calibration drift, etc.). Quality control has always been a task of great importance for missions that produce and use Earth observation sensor data on a daily routine basis. Now that a massive amount of sensor data is available – both historical and in the number of different sensors – the subject of automatic quality control and consequent corrections becomes only more of practical importance. The work of Szantai et al. (2011) which addresses anomalies in optical imagery of different geostationary satellite imagery is interesting. Our work complements this as it analyses the whole history of one family of geostationary sensors, i.e. the MFG satellites, for which the results are stored in a database that during later reprocessing can be consulted. Examples in the literature of image restoration – a process that normally follows after the quality assessment phase – include the correction for radiometric anomalies (Ruethrich et al., 2019; John et al., 2019) and the correction of the point spread function characteristics to compensate for sensor contamination (Doelling et al., 2015; Khlopenkov et al., 2015).

For the MFG satellites, EUMETSAT keeps an archive of the level 1.0 data (raw images with geolocation tie points but prior to rectification and calibration) and level 1.5 data (rectified to a fixed geolocation grid and calibrated). These data were archived in near-real time, meaning no corrections have been applied beyond that in the original processing. For a planned reprocessing of the level 1.0 to level 1.5 data addressing data anomalies, it is mandatory to create a consistent and as-complete-as-possible set of information concerning each Meteosat image, including its metadata and the detection and flagging of anomalies in the measurements. An image anomaly is defined as an anomaly in the radiometric content of an image not caused by the rectification process, a satellite manoeuvre, or scheduled change in satellite parameters (decontamination, gain changes). If such unexpected image radiometric anomalies occur, they can have a very detrimental impact on the use of the images. Except for problems such as wrong gain settings or wrong channel configuration, they will usually occur within the commissioning phase of a new satellite or in the period after just taking up operations with a new satellite. However, they still can occur suddenly on an operational satellite due to system failure. Another type of anomaly can be found in the metadata of the images. For various reasons, including operator error and control software problems, the metadata can be inconsistent or incomplete with respect to the scientific contents of the images. This paper tackles the issue of image anomalies in the MVIRI measurements by processing the whole archive of MVIRI images to detect and flag any anomalies present.

The paper is organized as follows. Section 2 briefly describes the MVIRI data and known anomalies in them, Sect. 3 presents the algorithms to detect and flag such anomalies, Sect. 4 discusses the results, and Sect. 5 provides conclusions and an outlook.

**Table 1.** List of satellite names, operational mission with nominal sub-satellite longitude position in brackets, the main years of operation (EUMETSAT, 2014), and the total number of level 1.0 data files in the archive for a given satellite.

Satellite	Mission	Main operational years	Number of level 1.0 files
Meteosat-2	0-degree (0°)	1981–1988	110 195
Meteosat-3	0-degree (0°)	1988–1991	89 614
Meteosat-3	ADC (50° W)	1991–1993	
Meteosat-3	XADC (75° W)	1993–1995	
Meteosat-4	0-degree (0°)	1989–1994	75 728
Meteosat-5	0-degree (0°)	1991–1997	210 260
Meteosat-5	0-degree RSS (0°)	1997–1997	
Meteosat-5	IODC (63° E)	1998–2007	
Meteosat-6	0-degree (0°)	1996–1998	144 773
Meteosat-6	0-degree RSS (0°)	2000–2007	
Meteosat-6	IODC (67° E)	2007–2009	
Meteosat-7	0-degree (0°)	1998–2006	329 334
Meteosat-7	IODC (57° E)	2006–2017	

**Table 2.** Spatial and spectral characteristics of MVIRI visible (VIS), thermal infrared (IR), and water vapour (WV) channels.

Channel	Res. nadir (km)	Nominal spectral band (µm)
VIS 0.7	2.25	0.40–1.10
WV 6.4	4.5	5.70–7.10
IR 11.5	4.5	10.5–12.5

## 2 Meteosat First Generation measurements

During the lifetime of the Meteosat First Generation series, the satellites were operated at different orbit locations. The primary 0° longitude orbit position has been covered from the start, supplemented by the so-called Indian Ocean Data Coverage service (IODC) from 1998 onwards with MFG satellites located at 57 and 63° E. IODC has been continued with the first Meteosat Second Generation (MSG) satellite operating from 41.5° E. Furthermore, in the first half of the 1990s, the Meteosat-3 satellite was moved to the western Atlantic to support coverage of the United States of America, thereby providing the so-called Atlantic Data Coverage (ADC) and Extended Atlantic Data Coverage (XADC) services from orbit positions at 50 and 75° W, respectively. Some of these satellites were also operated in rapid scanning mode (RSS) – the details are shown in Table 1.

The Meteosat Visible Infra-Red Imager (MVIRI) instrument measures in three spectral channels (Table 2): visible, water vapour (WV), and thermal IR. The VIS channel has

two detectors, VIS-South and VIS-North. There are 48 acquisition slots in a day (one every 30 min) and within one slot a MVIRI scan is created. A MVIRI scan consists of 3030 scan lines, where 2500 scan lines belong to the image-acquiring forward scan. The VIS-South and VIS-North detectors create 5000 samples along each scan line and the IR and WV detectors create 2500 samples along each scan line. It takes 25 min to complete the forward scan. Furthermore, the total size of all 1.0 files that were being processed is 47 TB.

### 3 Anomaly detection methods

Anomaly detection can be approached in various ways, ranging from machine-learning techniques to image processing in combination with classification techniques (Hodge and Austin, 2004). Each approach has its own merits, such as the effort to implement the anomaly detection system, the performance of the detection process, and the trust the end-users ultimately have in the approach selected. The implementation effort depends on the different types of anomalies one can expect in the dataset and whether it is possible to limit the number of anomaly types. If there is no good understanding of what types of anomaly one can expect, it will be difficult to program the algorithms that detect the presence of an anomaly in an image, but one can better concentrate on algorithms that define the nominal case in order to detect deviations from the norm. However, such an algorithm will be limited in its capability to classify or identify the type of anomalies. The performance of an anomaly detector can be described in terms of the probability of detection (PoD), the false alarm rate (FAR), and the specificity of the detection (i.e. whether the anomaly can be isolated to a subset of affected pixels rather than discarding the whole image). Finally, the trust that the ultimate end-users have depends primarily on how much one can understand the workings of the algorithm and whether there is some indication by the algorithm as to why an image is classified as anomalous (e.g. by providing an overview of the affected pixels and the type of anomaly).

A machine-learning approach is appealing from an implementation effort point of view, but in this case, the more classical approach of image processing using a manually selected array of anomaly classification algorithms has been used. The benefit of this well-known approach is that, during the investigation and development process, knowledge is generated on the exact appearance of the various anomaly types, which lead to improved end-user trust and a better chance of identifying the source of errors. There is also more control to tune the quality of the anomaly detection of the analysed image, especially in terms of specificity, and to prevent over-training of an algorithm on a limited dataset.

The remainder of this section describes the process of analysing a limited test set of data to identify anomalies and tune algorithms, and improving the quality of the analysis.

We call this subset of representative images the *training set*. It should be understood that this training set is used for visual inspection and to learn about the possible anomalies, and not for the automatic training of machine-learning algorithms. Next, several anomaly categories are presented and the development process of the detection algorithm will be discussed. Finally, to provide a feel for the types of algorithms applied, three anomaly detection concepts will be briefly described.

#### 3.1 Creating the training set

During the development of the MFG satellites and the years of operation, valuable knowledge on all kinds of topics related to the satellite and its application has been created. This knowledge can be related to typical low-level sensor aspects, such as signal-to-noise ratio or crosstalk, but also to the data processing of the sensed data and the applied corrections. For the MFG satellites, the active development period was several decades ago and the involved scientists and engineers are largely not available anymore. As a result, the present knowledge on anomalies in the dataset due to malfunctioning sensors or incorrect application is limited and diminishing with time. However, this lack of background information is also an opportunity to analyse the historical dataset from first principles instead of only focusing on known anomalies.

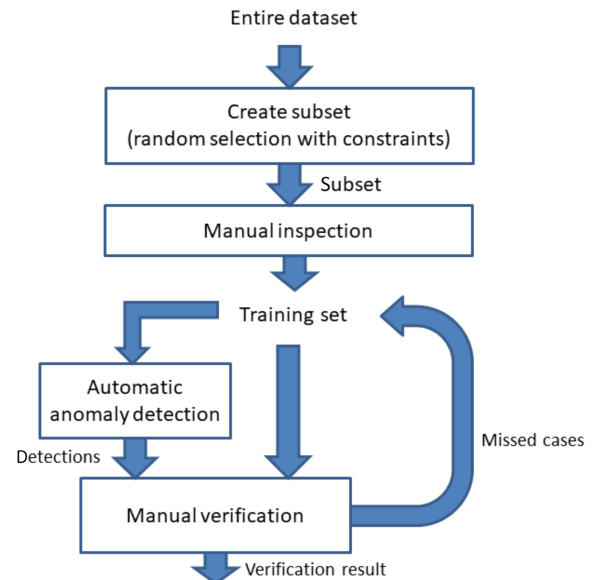
Therefore, the first step was a manual study of anomalies in the historical dataset. Due to the high number of images (around 1 million), only a limited subset can be manually examined. The objective was to find a representative set of anomaly types that one could expect in the whole dataset. Therefore, a training dataset was created with representative coverage of the relevant channels, periods, and satellites. Specifically, samples were randomly selected, but with several constraints in order to avoid large gaps in time or that a satellite, channel, or typical time slot (e.g. 1 January on 12:00 UTC) is over-represented. The dataset should be as small as possible for practicality reasons, but also large enough to cover all anomaly types. As most anomaly types (e.g. stray-light effect) will affect multiple channels in the same time slot, the dataset contains only one channel from any particular time slot, such that the size of the dataset is kept small while maximizing the number of independent images searched for anomalies. The main interest is in anomalies related to sensor failure or radiometric effects present in level 1.0 data, but it is also possible that the level 1.5 rectification process could introduce new anomalies, and thus the training-set dataset consists of both level 1.0 and level 1.5 images. This dataset was inspected manually to characterize

- types of anomalies,
- frequency of occurrence,
- appearance and severity of an anomaly, and
- origin of the anomaly/root cause.

Each image from the dataset is evaluated to determine whether it contains an anomaly. The dataset together with the evaluation is called the “training set” and is later used to tune and evaluate the performance of the automatic detection software. The manual inspection process requires several iterations to converge on consistent human decision criteria for flagging images. A difficult aspect is that the appearance of an anomaly in an image or for a particular pixel is modelled as an effect that is present or is not present (i.e. it cannot be partially present). In several cases, the severity of the appearance of the anomaly is small, and it is doubtful whether an image or a pixel should be flagged.

### 3.2 Improving the quality of the training set

The manual inspection of the MVIRI images is a difficult and time-consuming job. A dedicated tool has been developed to speed up and improve the quality of this manual inspection process. The inspection process also determines which anomaly types exist and what they look like, which is a learning process. To avoid inconsistent (human) judgements during this learning process, the images are inspected in several iterations. The manual inspection has been executed with the greatest care, but we also have to conclude that the quality of a human-inspected dataset is lower than desired. Common mistakes include inconsistent detection accuracy, mistakes concerning anomaly types, and missed pixels. On the other hand, humans are very good at detecting patterns and abnormalities in images, which would probably not be detected by any algorithm with very limited a priori information. To achieve the highest quality possible (no missed cases or false cases), an iterative strategy was chosen where the manual inspection and evaluation is corrected by algorithms (see Fig. 1). With this strategy, the training set is initially defined by manual inspection. New anomaly cases are detected by algorithms based on the initial evaluation and, after manual inspection and confirmation, the new cases are added to the training set. The algorithms used to detect new cases can be all kinds of detection algorithms, but include the to-be-developed anomaly detection algorithms. In our case, the algorithms used to correct the manual inspection were a combination of the final automatic detection algorithms and other (generic) detection algorithms. An example of a generic detection algorithm is one that detects whether the average intensity and the standard deviation are within a certain range. The usage of a detection algorithm is especially useful when the appearance of a particular anomaly cannot be visually detected by a human in a single image. The accuracy of such generic detection algorithms can be quite low, but they help with flagging images where the severity of the anomaly is quite low. An example in the literature of such a generic detection algorithm in use was to detect cases of the “loose cold optics” issue of the Meteosat-6 satellite (Holmlund, 2005). The proposed method has low detection accuracy but, with manual inspection and evaluation of the algorithm’s internal



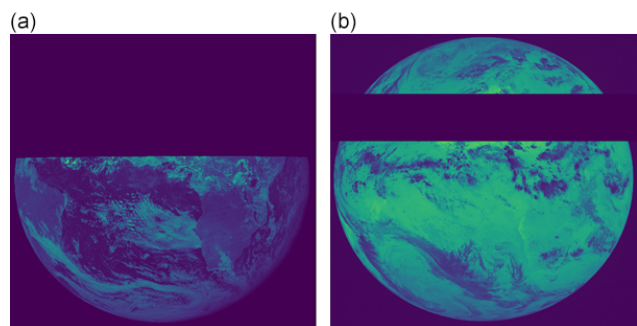
**Figure 1.** Approach for creating a training set, where a manually inspected image is corrected by algorithms.

calculations, the quality of the training set can be increased. The strategy of improving the quality of the dataset is applied during the entire development of the anomaly detection algorithms and during the evaluation of the detection performance (see also the algorithm description as provided in the Supplement to this article).

### 3.3 Anomaly types

The training-set dataset was analysed and 30 different anomaly types were defined. The specifics of the defined anomaly types and their appearance will be unique to the MFG satellites, but the general anomaly origin or category will also hold for other similar geostationary imaging satellite types, such as GOES (Schmetz and Menzel, 2015; Considine, 2006) and GMS/MTSAT (Tabata et al., 2019). For the MFG anomalies, some examples of categories follow (see Table 3 for a complete overview).

- Missing or corrupt data (see Fig. 2). All pixels of an image or a scan line have the value “0” or an obviously incorrect value.
- Low-quality sensory data (see Fig. 3. For example, the signal-to-noise ratio of the image is much lower than expected or the pixels are affected by a disturbance source.
- Unexpected behaviour of (historical) processing. The processing software used has changed during the lifetime of the satellites, and these changes sometimes resulted in different behaviour. An example of unexpected behaviour is a different definition of the start time of a scan or that pixels have been set to “0” for various different reasons.



**Figure 2.** Examples of missing data. Taken from file METEOSAT3-MVIRI-MTP15-NA-NA-19900816133000 and channel WV (a) and file METEOSAT2-MVIRI-MTP10-NA-NA-19840215160000 and channel IR (b). The pixel values in this figure are the raw digital counts and are mapped using the viridis colour map with 0 represented by dark blue and 255 by yellow (© EUMETSAT).

- Stray-light-related anomalies (see Fig. 4). Indirect illumination of a light detector by internal reflections; e.g. in the right locations, the Sun will reflect off internal components of the telescope and onto pixels that are not looking at the Sun. This stray-light effect will directly affect the various detectors (VIS, IR, and WV), but it can also indirectly affect the consecutive scans for reasons which are currently not known to us. With the MFG satellites, the WV images were affected several hours after the initial stray-light effect. Images have also been affected by stray light from the Moon.
- Unstable optics-related anomalies. The Meteosat-5 and -6 satellites suffer from known hardware issues related to the optics, which has the effect that the sensitivity changes in time.

The training set also contains level 1.5 images, but we have not discovered any anomaly that is related to the level 1.5 rectification process. In general, it holds that discovered anomalies in the level 1.5 images are better recognizable in the level 1.0 images. The rectification process (Wolff, 1985) blurs anomalies on the pixel level, so that if an anomaly only affects a single scan line in the level 1.0 images, its effect on the level 1.5 image will be a blurry curved line.

Table 3 shows an overview of the defined anomaly types.

### 3.4 Development strategy anomaly detection algorithms

The creation of the training set gave insight into the appearance and the probability of occurrence of the defined anomaly types. Some anomaly types occur very often, while others only had a few examples in the training set. Some anomaly types are limited to a particular satellite and others are related to a particular channel or period of the day. The training set covers more than 2500 time slots, but still is far too small to calculate reliable probabilities of occurrence for

several anomaly types. Therefore, during the development of the detection algorithms, this uncertainty in the probability of occurrence must be continually taken into account. For each anomaly type, the most likely root cause or origin of the anomaly is determined to avoid mistakes in the estimation of the probability of occurrence in a certain situation (channel, satellite, period of the day, etc.).

Distinctive features for specific anomalies can be various metadata parameters, such as satellite id or time, but in general the focus is on image-based parameters. The following data sources can be used for the detection of an anomaly.

- Metadata parameters of the file (satellite id, date, time, geo-location, etc.)
- Image data of a channel
- Other channels of same time slot
- Series of consecutive images

In general, it is preferable to minimize the required number of parameters examined for an anomaly because each parameter can be affected by other anomalies than the targeted one. Therefore, when multiple parameters are used, care must be taken to ensure that each parameter is genuinely necessary. Also, feature calculations are complicated by the fact that files (or parameters) may be missing; this mainly occurred when the majority of a scan is affected by the stray-light effect. In general, anomalies are more recognizable in the raw data (level 1.0 files), so a choice was made to do anomaly detection only on level 1.0 files. If the detection requires a series of consecutive images, the algorithm has to perform some kind of registration between the images. The level 1.0 file contains information for the alignment or registration, which is used for an approximated rectification based on a linear homography transform (transform is defined by a  $3 \times 3$  matrix multiplication). The (internal) approximated rectification results are verified on the basis of a cross-correlation. If a “better” shift ( $X$ ,  $Y$  displacement) between images can be found with cross-correlation, this improved shift is applied to register the images.

In addition to detecting the presence of an anomaly, the area it affects in the image needs to be determined. The affected area can be described on the following levels.

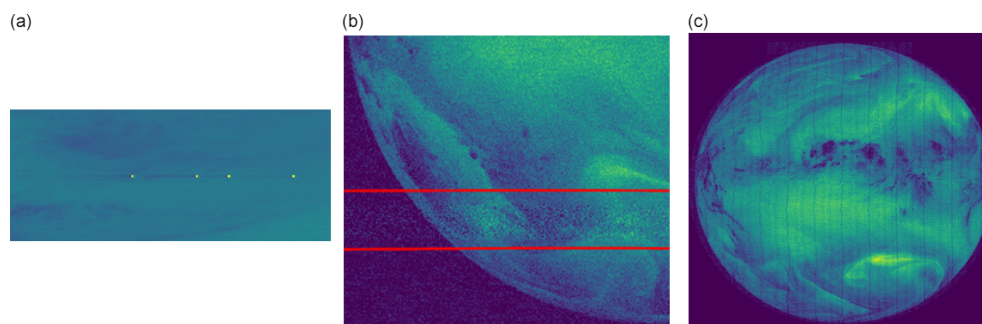
- Image level, where the entire or majority of the image is affected by an anomaly.
- Scan-line level, where a single or multiple scan lines are affected by an anomaly.
- Pixel level, where the anomaly affects a pixel or multiple pixels.

In general, the affected area was already determined by the calculation of the distinctive features for anomaly detection. The affected area is stored in a database and, to keep

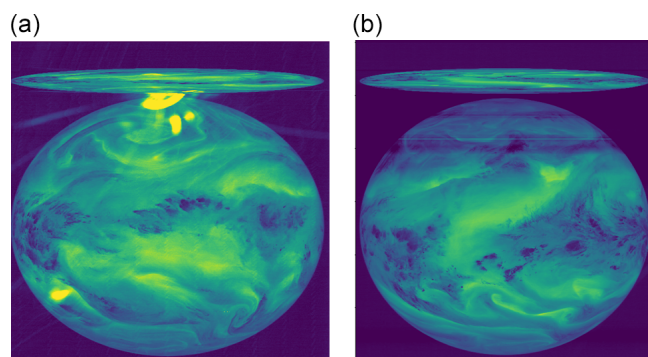
**Table 3.** Description of the defined anomaly types in the MFG measurement and the affected channels. For the channels, *all* means the anomaly affects all channels simultaneously, whereas *each* means that each channel is analysed individually.

CATEGORY	TYPE	Channels	Description
artefact	misalignment	all	Scan lines are not aligned properly and the east–west horizon of the Earth is not a continuous curve.
	over-illumination (overflow)	each	Over-illuminated pixels have an incorrect value of 124 instead of the maximum value of 255.
	tilted line	WV	In an image, a line under an angle 19° (from the vertical) is visible.
celestial body	celestial body: the Moon	all	The Moon is present in the MVIRI image.
	celestial body: undefined	all	An artefact appearing similar to a celestial body was detected in the space area of the image, but the Moon can be excluded due to orbital position
corrupt or missing	completely black	each	All or almost all pixels have an intensity lower than 10 (a threshold significantly higher than background noise).
	corrupt file	all	The size of the level 0 file is too small to contain data of all scan lines.
	hanging scan line	all	Position of detector has not changed
	incomplete image	each	Forward scan time/length is too small to capture the entire Earth. Note that if a rapid scan image is analysed, the captured image is considered as incomplete because the software is currently only designed for full Earth images.
	invalid signal	each	According to the metadata, a channel is invalid.
	large black area	each	The image contains a large black area, where several scan lines are completely black (intensity is 0).
	large white area	each	The image contains a large white area, where several scan lines are completely white (intensity is 255).
	no sub-images	each	The scan does not contain any forward scan.
hot pixel	hot pixel pattern 1	WV	A typical pixel pattern on a single scan line.
	hot pixel pattern 2	all	A typical pixel pattern on a single scan line.
	hot pixel pattern independent	each	Randomly distributed hot pixels (high intensity and very unlikely compared to the neighbouring pixels).
unstable optics	unstable optics	WV and IR	The observed sensitivity of the detector is not constant in time. This anomaly mainly appears with the known optical hardware issues of Meteosat-5 and 6.
low SNR	low SNR: scan line	WV	The observed noise level in a scan line is much higher than the observed noise level in the entire image.
metadata	EFF position corrupt	all	The stored position of the satellite in the Earth Fixed Frame format is corrupt.
	orbit position empty	all	The stored position of the satellite is empty.
	parameter empty	all	A metadata parameter, which should have a value, is empty or zero.
	start time: forward scan	all	The start time definition used is unexpected. Here, the start time of the “scan” is equal to start time of the forward scan.
	start time: southern horizon	all	The start time definition used is unexpected. Here, the start time of the “scan” is equal to the moment when the southern horizon is detected.
	start time: start image	all	The start time definition used is unexpected. Here, the start time of the “scan” is equal to the start time of the entire scan.
	start time: undefined	all	The definition of the start time cannot be determined.
	value unexpected	all	A metadata parameter has an unexpected value (outside a certain range).
raw data manipulated	background noise removed	each	Raw data has been changed and pixels that should contain background noise have “0” as their value.
	background noise removed and noise added	each	Raw data has been changed. Besides the removal of the background noise, the intensity of all pixels might have been adjusted.
	the number of scan lines changed	each	The number of valid scan lines has been changed.
stray light	direct stray light	each	Image is affected by parasitic light via indirect optical path, which results in a typical pattern where the pixels have a higher intensity.
	indirect stray light	WV	After the appearance of the direct stray light effect, several images can be affected. With the affected images, several scan lines will have a significant lower intensity. The affected scan lines with the indirect stray light anomaly are the scan lines, where in a preceding image the direct stray-light effect was present.
	reflection of the Moon	WV	An over-illuminated, fingernail-/crescent-shaped blob. It often appears on the right side of the Moon.
suspicious pattern	suspicious pattern	WV	The image contains a non-physical pattern.

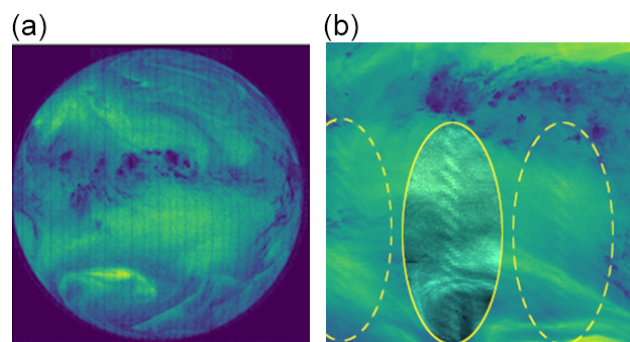




**Figure 3.** Examples of low-quality sensory data. (a) A scan line contains an interference pattern (file: METEOSAT4-MVIRI-MTP10-NA-NA-19911216110000, channel: WV). (b) A block of scan lines (inside the red lines) with a much lower signal-to-noise ratio than neighbouring scan lines (file: METEOSAT2-MVIRI-MTP10-NA-NA-19831016120000, channel: WV). (c) Image contains an unknown disturbance pattern (file: METEOSAT2-MVIRI-MTP10-NA-NA-19810817223000, channel: WV). The pixel values are the raw digital counts (see also Fig. 2; © EUMETSAT).



**Figure 4.** (a) Image affected by a stray-light anomaly (file: METEOSAT4-MVIRI-MTP10-NA-NA-19900415003000, channel: WV). (b) Several hours later, still showing follow-on effects after the stray-light anomaly (file: METEOSAT5-MVIRI-MTP10-NA-NA-20060317213000, channel: WV). The pixel values are the raw digital counts (see also Fig. 2; © EUMETSAT).



**Figure 5.** Two examples of the “suspicious pattern” anomaly: in (a) the vertical stripes are visible (file: METEOSAT2-MVIRI-MTP10-NA-NA-19810817223000, channel: WV). In (b) the ovals indicate the position of the suspicious pattern; the middle oval has been altered to enhance the visibility of the pattern (file: METEOSAT4-MVIRI-MTP10-NA-NA-19901015013000, channel: WV). The pixel values are the raw digital counts (see also Fig. 2; © EUMETSAT).

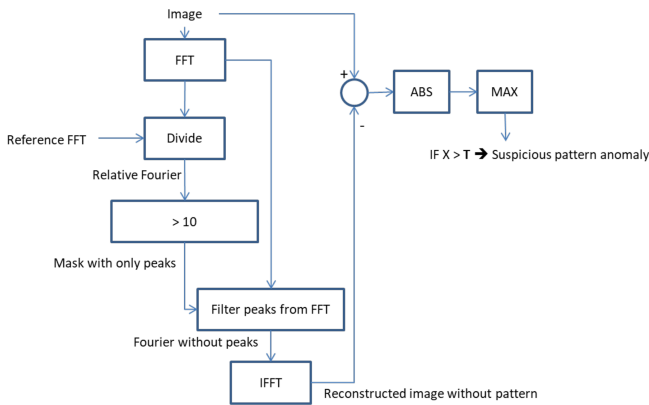
### 3.5 Example complex-anomaly detection: “suspicious pattern”

the database efficient, the affected area (scan line or pixels) is approximated by a list of rectangles (each specified by  $X$  and  $Y$  coordinates of a corner plus  $X$  and  $Y$  size).

For each anomaly type, a dedicated algorithm needs to be developed. The algorithm development for some anomaly types, such as the missing data anomaly (several scan lines are missing), is quite straightforward. Others, such as the more image-processing-based anomalies, can be very challenging and require elaborated, innovative detection concepts. The next sections will briefly describe the detection concept of three difficult anomaly types, to illustrate the variety of processing steps used. It will mainly focus on the basic detection concept and not on all important pre-processing steps and details to reach robustness under all circumstances.

Figure 5 shows two examples that suffer from the “suspicious pattern” anomaly. This anomaly type may have various appearances (due to unknown root causes), which all result in a suspicious repeating pattern. The repetitive pattern is clearly visible, but the magnitude is still quite small. Such repetitive patterns can be detected by the analysis of the 2-D FFT spectrum, where this anomaly will introduce peaks in the 2-D FFT spectrum. As the variation in appearance is very large (magnitude, spatial separation of the repeats, vertical or horizontal pattern), we do not search for a particular pattern/peak, but compare the observed 2-D FFT spectrum with the expected 2-D FFT spectrum.

To detect these peaks in the 2-D FFT spectrum, we divide the observed 2-D FFT from a single image by the expected



**Figure 6.** Conceptual processing pipeline for detecting the suspicious pattern anomaly.

2-D FFT from the particular satellite. The expected 2-D FFT from a satellite is calculated by averaging the 2-D FFT from 100 images that did not contain any anomaly. If the ratio between the two 2-D FFT spectra is larger than a threshold (count value of 10), we define it as a peak. The threshold has been manually determined with the aim of detecting peaks in images where they would be identified also by human eye.

A (clear) peak in the 2-D FFT spectrum does not always result in a noticeable (by humans) pattern in the spatial domain (normal image). This especially holds for peaks that correspond to fast changing patterns with a small magnitude (smaller than  $1/255$  of the maximum intensity). The effect of the detected FFT peaks can be calculated by comparing the difference between the original and reconstructed images. The reconstructed image is calculated by the inverse FFT of the 2-D FFT spectrum with the peaks removed/reduced to a normal value. Only if this difference exceeds a threshold ( $T$ ) will an anomaly be flagged. The flowchart of Fig. 6 describes the process of detecting suspicious patterns in a schematic fashion.

### 3.6 Example complex-anomaly detection: “direct stray light”

The design of the MFG satellites suffers from the issue that the detector can be illuminated via an indirect optical path, which results in the occurrence of the so-called direct stray-light anomaly. This “parasitic” light causes a pattern with an increased intensity in the image. The observed pattern often contains characteristic bows or arcs, but its appearance is a little bit different in every instance. Figure 7 shows three example images affected by the direct stray-light anomaly.

In a series of images, the stray-light pattern moves across the scene quite quickly compared to the normal movements of the background (clouds, etc.), which enables its detection. Figure 8 shows three sequential images that are affected by the same direct stray-light anomaly. For detection, we as-

sume that an affected pixel is (significantly) brighter than the same pixel of the preceding and consecutive images.

To be able to use this assumption, it is essential that the raw images are aligned with each other and that no other anomalies have affected them. This allows the algorithm to determine whether a pixel is affected by this anomaly (see Fig. 9). Note that the bow in the lower left corner of the image has not been identified. The reason for this is that the algorithm compares the current image with the previous one. In this case the previous image also contained the same bow, and therefore the algorithm fails to identify the bow as anomalous.

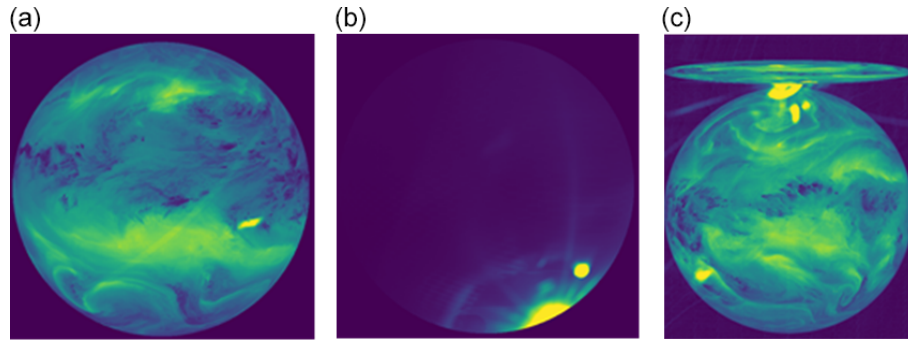
### 3.7 Example complex-anomaly detection: “unstable optics”

Meteosat-5 and -6 both suffer from a known optical hardware issue, which affects the sensitivity of the detector (Koepken, 2004). The sensitivity of the detector can vary by just a few percent and the effect is not noticeable by a human, which makes it hard to manually select images that are affected by the hardware issue. The detector’s sensitivity is not continuously affected by these issues and the appearance and effect magnitude changes from time to time. Koepken (2004) has shown that the occurrence of this anomaly can be detected by analysing the average IR intensity (over an entire image) over time or by cross-referencing with another satellite. However, requiring a complete image average means there is only one observation every 30 min, while the anomaly can continuously change and, in the meantime, the temperature of the Earth also changes (e.g. as night progresses across the Earth from the satellite’s viewpoint). Therefore, this method for detection is not very sensitive or reliable. Cross-referencing measurements from the affected satellite with data from another satellite is unfortunately not always possible. It is also preferable to have a detection concept that is independent of other data sources. Therefore, the selected detection approach models the effect of the anomaly as an (intensity) offset per scan line per image, calculated by least-squares optimization. The optimization makes two assumptions: (i) the average intensity of a scan line in an image is equal to the average intensity of the same scan line in the preceding image; (ii) the effects of the anomaly average out over  $N$  time slots. The first assumption is in general valid for registered images without any anomaly. The second assumption is a direct consequence of a time-dependent stochastic process. Based on the results, we can conclude that the second assumption is valid for the known optical hardware issues of Meteosat-5 and -6.

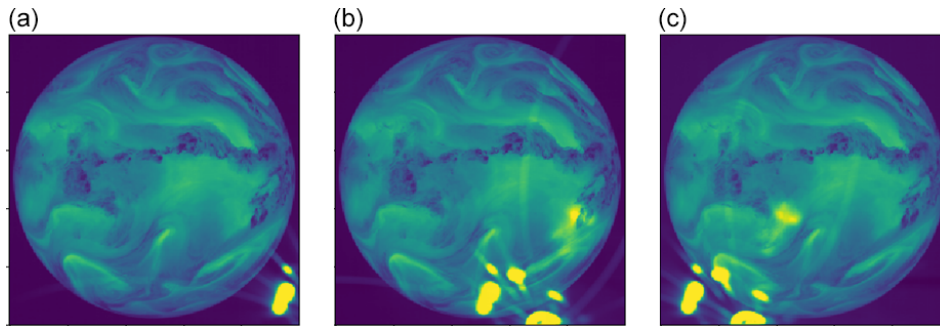
The bias for each scan line per image corresponding to this anomaly type can be calculated by solving a linear system of equations, from which we define the following variables:

- $X[i, j]$  = measured average intensity of scan line  $j$  from image  $i$ .

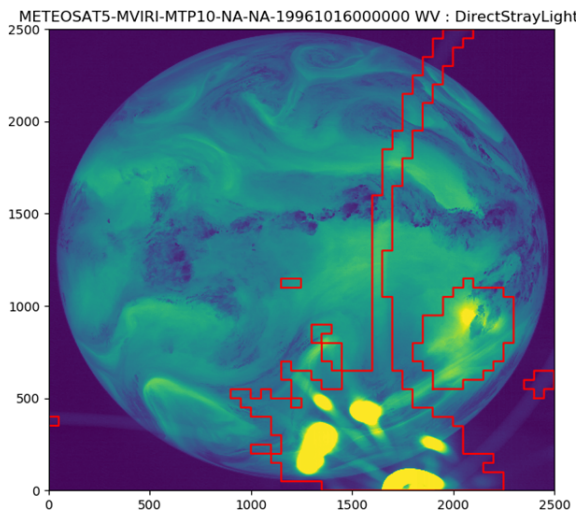




**Figure 7.** Three examples of the direct stray-light effect anomaly. **(a)** Taken from file METEOSAT7-MVIRI-MTP15-NA-NA-20130717210000 and channel WV. **(b)** Taken from file METEOSAT7-MVIRI-MTP15-NA-NA-19991016003000 and channel VIS. **(c)** Taken from file METEOSAT4-MVIRI-MTP10-NA-NA-19900415003000 and channel WV. The pixel values are the raw digital counts (see also Fig. 2; © EUMETSAT).



**Figure 8.** Three sequential images affected by the direct stray-light anomaly. **(a)** Taken from file METEOSAT5-MVIRI-MTP10-NA-NA19961015300000 and channel WV. **(b)** Taken from file METEOSAT5-MVIRI-MTP10-NA-NA19961016000000 and channel WV. **(c)** Taken from file METEOSAT5-MVIRI-MTP10-NA-NA19961016300000 and channel WV. The pixel values are the raw digital counts (see also Fig. 2; © EUMETSAT).



**Figure 9.** Segmentation result of the direct stray-light anomaly detection. Taken from METEOSAT5-MVIRI-MTP10-NA-NA19961016000000 and channel WV. The pixel values are the raw digital counts (see also Fig. 2; © EUMETSAT).

–  $B[i, j]$  = bias due to the unstable optics anomaly of scan line  $j$  from image  $i$ .

Assumption 1 results in the following equation:

$$X[i, j] - B[i, j] \approx X[i - 1, j] - B[i - 1, j] \Rightarrow X[i, j] - X[i - 1, j] \approx B[i, j] - B[i - 1, j].$$

Of course, these equations hold for every image  $i$ . Assumption 2 results in the following equation:

$$B[i, j] + B[i + 1, j] + B[i + 2, j] + \dots + B[i + N, j] \approx 0.$$

The equations for all images and scan lines can be stored in a matrix. An example of a part of the matrix, where the anomaly is averaged out over five images with weight  $W$ , is

$$\begin{pmatrix} -1 & 1 & & & & & & \\ & -1 & 1 & & & & & \\ & & -1 & 1 & & & & \\ & & & -1 & 1 & & & \\ & & & & -1 & 1 & & \\ W & W & W & W & W & & & \\ & W & W & W & W & W & & \\ & & W & W & W & W & W & \end{pmatrix} \cdot \begin{pmatrix} B[i, j] \\ B[i+1, j] \\ B[i+2, j] \\ B[i+3, j] \\ B[i+4, j] \\ B[i+5, j] \\ B[i+6, j] \\ B[i+7, j] \end{pmatrix} \approx \begin{pmatrix} X[i+1, j] & - & X[i, j] \\ X[i+2, j] & - & X[i+1, j] \\ X[i+3, j] & - & X[i+2, j] \\ X[i+4, j] & - & X[i+3, j] \\ X[i+5, j] & - & X[i+4, j] \\ X[i+6, j] & - & X[i+5, j] \\ X[i+7, j] & - & X[i+6, j] \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (1)$$

In our case, the linear system covers in total 21 consecutive time slots, where the anomaly averages out over 5 time slots and the anomaly bias is modelled per 100 scan lines. The calculated bias per image per scan line can be stored as an array (like Fig. 10b) to better see how bias per scan lines changes within an image or in time.

The calculated bias of images that have been affected by this anomaly will be, in general, close to one digital count. The detection of this anomaly in an image is based on the average magnitude of the calculated scan lines' biases in a particular image (see also the algorithm description as provided in the Supplement to this article).

## 4 Results

For each anomaly type, a dedicated algorithm has been developed. The detection performance of the algorithm on the training set (covering 2500 time slots) has been manually verified. During this verification process, we noticed that algorithms found more anomalies in the training set than were manually found. After inspection of the new detections, it appeared that in most cases the algorithms were correct. If we look at the overall detection performance of the algorithms on the basis of the training-set dataset, 97.7 % (2.3 % missed cases) of the anomalies are successfully (true positive) detected and 2.7 % of the detections are incorrect (false positive). The overall probability of detection of the anomalies has been determined based on the results from the training set. The POD of each anomaly type has not been determined, because the number of occurrences in the training set is considered to be too low to reliably calculate the POD. For some anomaly types the sensitivity level of the algorithm (should an image where an anomaly is vaguely visible be flagged or not) could be subject to end-user's preference, and as such the sensitivity level will affect the POD and the false alarm

rate of the algorithm. For simple anomaly categories, such as "corrupt or missing" and "hot pixel", the anomalies are detected in all cases. For the more complex anomaly types, such as "direct stray light" and "suspicious spectrum", the POD is around 90 %.

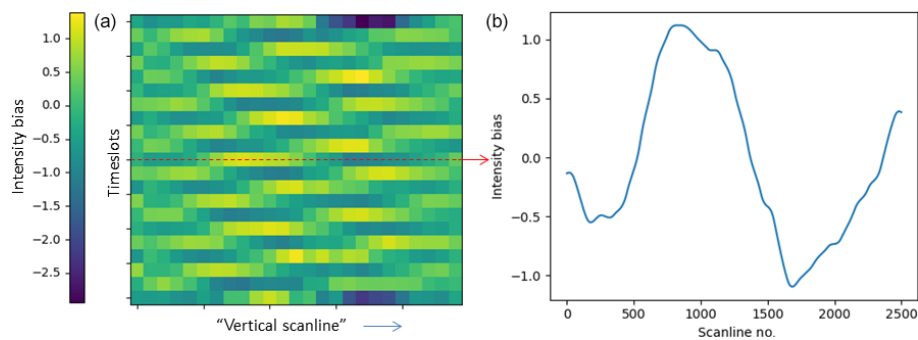
After the verification of the training set, all images from the MFG satellites have been processed and the anomaly detections stored in a separate database. An overview of the detected anomalies in the entire MFG dataset is presented in Table 4. The percentage of level 1.0 files that contain an anomaly of a certain type (affecting any of the channels) is shown separately per satellite. The MET6 dataset contains mostly RSS images, which explains some of the high anomaly rates.

The database can be used to create statistical analyses about the anomaly distribution or to filter images (or even pixels) for reprocessing campaigns or for specific and sensitive use cases such as cross-calibration.

## 5 Discussion and conclusions

Monitoring the quality of input data is of great importance in guaranteeing the correctness of the results of any analysis. For long-term studies, such as those on the Earth's climate, there is a need to combine datasets of various instruments, including those of early sensors that were not originally well quality-controlled. Later efforts to identify anomalies face multiple difficulties – loss of the original human expertise, limited documentation, and datasets too large to assess manually in retrospect. A practical quality-assessment system must be based on automatic means and, rather than merely removing imperfect data from a period where observations are limited, must be able to support a wide variety of future uses by supplying detailed and precise information on the form and impact of anomalies. In addition, a uniform approach towards assessing the quality of data products is of great benefit to improving consistency over multiple sensors.

This paper describes a general method to screen an EO image database with a cumulative observation history of approximately 40 years. It has been shown that the method of using dedicated anomaly detection algorithms is sufficiently powerful to detect a wide array of anomalies, ranging from clear faults to subtle problems related to stray light that occur only under certain celestial constellations. The main challenge was to develop the methods such that the algorithms accurately detect the images that are affected by the anomalies and, within the images, which areas are affected. With respect to the first objective, the probability of detection for affected images has been established at 97.7 % and the false alarm rate of the method is 2.7 %. The specificity within an affected image of the method is subjectively very good, and most of the detection algorithms are able to highlight only those pixels that are affected.



**Figure 10.** Example of the Meteosat-5 optical hardware issue in time, where the image (a) shows the anomaly's bias for consecutive images across the scan lines. The image (b) is for a particular time slot and shows the anomaly's bias across the scan lines. Taken from METEOSAT5-MVIRI-MTP10-NA-NA-19960515120000 and channel IR (© EUMETSAT).

**Table 4.** Percentage of level 1.0 files containing anomalies in the images and metadata.

Category	Type	MET2	MET3	MET4	MET5	MET6	MET7
artefact	east–west horizon misaligned	0.1	< 0.1	0.1	< 0.1	< 0.1	< 0.1
	over-illumination	32.9	34.9	< 0.1	< 0.1	< 0.1	< 0.1
	tilted line	< 0.1	< 0.1	3.8	3.9	< 0.1	< 0.1
celestial body	celestial body: the Moon	0.3	0.4	0.3	0.4	0.2	0.4
	celestial body: undefined	0.1	0.1	0.1	< 0.1	< 0.1	0.1
corrupt or missing	completely black	3.4	1.1	0.6	< 0.1	2.5	0.4
	corrupt file	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
	hanging scan line	1.7	2.6	0.9	0.9	0.9	0.6
	hanging scan line: no sub-images	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
	incomplete image	1.4	2.5	1.4	0.2	62.4	0.5
	invalid signal	100	100	0.1	0.1	< 0.1	0.2
	large black area	3.8	4.0	2.7	2.8	2.3	1.1
	large white area	0.8	0.3	< 0.1	< 0.1	< 0.1	< 0.1
	no sub-images	0.2	0.4	0.8	0.3	1.0	1.1
hot pixel	hot pixel pattern 1	< 0.1	< 0.1	91.4	92.0	74.8	90.9
	hot pixel pattern 2	0.4	0.2	4.7	13.7	2.9	5.0
	hot pixel pattern independent	40.6	5.0	18.1	24.2	2.3	9.4
unstable optics	unstable optics	0.4	0.3	0.2	2.7	10.2	0.2
low SNR	low SNR: scan line	16.2	1.3	0.3	0.1	< 0.1	0.1
raw data manipulated	background noise removed	98.8	98.6	13.6	< 0.1	0.1	< 0.1
	background noise removed/noise added	91.1	10.9	0.7	< 0.1	0.4	< 0.1
	the number of scan lines changed	2.1	0.2	< 0.1	0.1	< 0.1	< 0.1
stray light	direct stray light	3.9	5.6	4.8	6.8	1.9	4.7
	indirect stray light	6.1	4.7	4.9	6.4	2.2	6.1
suspicious spectrum	reflection of the Moon	0.1	0.2	< 0.1	< 0.1	< 0.1	< 0.1
	suspicious spectrum	6.9	44.9	60.4	2.2	0.1	1.8
metadata	EFF position corrupt	26.6	81.8	1.7	0.2	< 0.1	< 0.1
	orbit position empty	2.3	1.3	2.0	0.3	1.1	1.5
	parameter empty	100	100	100	17.3	63.9	1.8
	start time: forward scan	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
	start time: southern horizon	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
	start time: start image	70.0	0.2	0.1	0.4	62.7	0.3
	start time: undefined	< 0.1	< 0.1	< 0.1	0.1	< 0.1	0.1
	value unexpected	100	100	100	17.3	63.9	1.8

The anomaly detection results for the full dataset of EUMETSAT's MFG satellites are stored in a dedicated database that can be consulted to better understand the distribution of anomalies over the complete dataset and to filter the image data so that long-term analyses are being conducted on quality-controlled input data.

The anomaly detection system will be an essential part of the quality-control system in future reprocessing and analysis work, and strengthens EUMETSAT's stewardship of the full MFG data archive by providing a consistent and data-based methodology for quality assessment. Although the anomaly detection algorithms have been tested on MFG data, it is believed that the approach can be used for other similar geostationary satellite instruments as well, such as those on MSG (Schmetz and Menzel, 2015), GMS/MTSAT (Tabata et al., 2019), and GOES (Considine, 2006), as no satellite-specific knowledge is needed to parameterize the detection algorithms.

**Data availability.** MFG level 1.0 data can be requested through the EUMETSAT help desk at ops@eumetsat.int.

**Supplement.** The supplement related to this article is available online at: <https://doi.org/10.5194/amt-13-1167-2020-supplement>.

**Author contributions.** Conceptualization was by VOJ, AB, FL, and JS; formal analysis was done by FL; investigation was conducted by FL; the methodology was by FL; project administration was by SL; software were provided by FL, PWGB, FR, and MGG; validation was done by JO and FL. Visualization was by FL; supervision was by JS and AB; FL, VOJ, and AB wrote the original draft; FL, AB, VJ, FR, JO, and MGG were involved in writing, reviewing, and editing.

**Competing interests.** The authors declare that they have no conflict of interest.

**Review statement.** This paper was edited by Andrew Sayer and reviewed by Jörg Trentmann and one anonymous referee.

## References

- Brogniez, H., Roca, R., and Picon, L.: A clear-sky radiance archive from Meteosat "water vapor" observations, *J. Geophys. Res.*, 111, D21109, <https://doi.org/10.1029/2006JD007238>, 2006.
- Considine, G. D. (Ed.): Geostationary Operational Environmental Satellite (GOES), chap. Eclipse, in: Van Nostrand's Scientific Encyclopedia, American Cancer Society, <https://doi.org/10.1002/0471743984.vse8611>, 2006.
- Doelling, D. R., Khlopenkov, K. V., Okuyama, A., Haney, C. O., Gopalan, A., Scarino, B. R., Nordeen, M., Bhatt, R., and Avey, L.: MTSAT-1R Visible Imager Point Spread Correction Function, Part I: The Need for, Validation of, and Calibration With, *IEEE T. Geosci. Remote*, 53, 1513–1526, <https://doi.org/10.1109/TGRS.2014.2344678>, 2015.
- Duguay-Tetzlaff, A., Bento, V. A., Göttsche, F. M., Stöckli, R., Martins, J. P. A., Trigo, I., Olesen, F., Bojanowski, J. S., Da Camara, C., and Kunz, H.: Meteosat Land Surface Temperature Climate Data Record: Achievable Accuracy and Potential Uncertainties, *Remote Sens.*, 7, 13139–13156, <https://doi.org/10.3390/rs71013139>, 2015.
- EUMETSAT: EUM/OPS/DOC/08/4698, available at: [https://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET\\_FILE&dDocName=PDF\\_METEOSAT\\_PRIME\\_SATELLITES&RevisionSelectionMethod=LatestReleased&Rendition=Web](https://www.eumetsat.int/website/wcm/idc/idcplg?IdcService=GET_FILE&dDocName=PDF_METEOSAT_PRIME_SATELLITES&RevisionSelectionMethod=LatestReleased&Rendition=Web) (last access: 5 February 2020), 2014.
- Hodge, V. J. and Austin, J.: A Survey of Outlier Detection Methodologies, *Artif. Intell. Rev.*, 22, 85–126, <https://doi.org/10.1007/s10462-004-4304-y>, 2004.
- Holmlund, K.: Satellite instrument calibration issues: Geostationary platforms, EUMETSAT NWP-SAF Workshop, available at: <https://www.ecmwf.int/node/15839> (last access: 5 February 2020), 2005.
- John, V. O. and Soden, B. J.: Temperature and humidity biases in global climate models and their impact on climate feedbacks, *Geophys. Res. Lett.*, 34, L18704, <https://doi.org/10.1029/2007GL030429>, 2007.
- John, V. O., Tabata, T., Ruethrich, F., Roebeling, R. A., Hewison, T., Stoeckli, and Schulz, J.: On the Methods for Recalibrating Geostationary Longwave Channels Using Polar Orbiting Infrared Sounders, *Remote Sens.*, 11, 1171, <https://doi.org/10.3390/rs11101171>, 2019.
- Khlopenkov, K. V., Doelling, D. R., and Okuyama, A.: MTSAT-1R Visible Imager Point Spread Function Correction, Part II: Theory, *IEEE T. Geosci. Remote*, 53, 1504–1512, <https://doi.org/10.1109/TGRS.2014.2344627>, 2015.
- Koepken, C.: Solar Stray Light Effects in Meteosat Radiances Observed and Quantified Using Operational Data Monitoring at ECMWF, *J. Appl. Meteorol.*, 43, 28–37, [https://doi.org/10.1175/1520-0450\(2004\)043<0028:SLEIM>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0028:SLEIM>2.0.CO;2), 2004.
- Mueller, R., Pfeifroth, U., Traeger-Chatterjee, C., Trentmann, J., and Cremer, R.: Digging the METEOSAT Treasure – Decades of Solar Surface Radiation, *Remote Sens.*, 7, 8067–8101, <https://doi.org/10.3390/rs70608067>, 2015.
- Rossow, W. B. and Schiffer, R. A.: Advances in Understanding Clouds from ISCCP, *B. Am. Meteorol. Soc.*, 80, 2261–2288, [https://doi.org/10.1175/1520-0477\(1999\)080<2261:AIUCFI>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2), 1999.
- Ruethrich, F., John, V. O., Roebeling, R. A., Quast, R., Govaerts, Y., Wooliams, E., and Schulz, J.: Climate Data Records from Meteosat First Generation Part III: Recalibration and Uncertainty Tracing of the Visible Channel on Meteosat-2–7 Using Reconstructed, Spectrally Changing Response Functions, *Remote Sens.*, 11, 1165, <https://doi.org/10.3390/rs11101165>, 2019.
- Schmetz, J. and Menzel, W. P.: A Look at the Evolution of Meteorological Satellites: Advancing Capabilities and Meeting User Requirements, *Weather Clim. Soc.*, 7, 309–320, <https://doi.org/10.1175/WCAS-D-15-0017.1>, 2015.

- Soden, B. J. and Bretherton, F. P.: Upper tropospheric relative humidity from the GOES 6.7  $\mu\text{m}$  channel: Method and climatology for July 1987, *J. Geophys. Res.*, 98, 16669–16688, <https://doi.org/10.1029/93JD01283>, 1993.
- Stoeckli, R., Bojanowski, J. S., John, V. O., Duguay-Tetzlaff, A., Bourgeois, Q., Schulz, J., and Hollmann, R.: Cloud Detection with Historical Geostationary Satellite Sensors for Climate Applications, *Remote Sens.*, 11, 1052, <https://doi.org/10.3390/rs11091052>, 2019.
- Szantai, A., Six, B., Cloché, S., and Sèze, G.: MTTM Megha-Tropiques Technical Memorandum, Quality of Geostationary Satellite Images, Tech. rep., SATMOS centre (Service d'Archivage et de Traitement Météorologique des Observations Satellitaires, available at: <http://meghatropiques.ipsl.polytechnique.fr/megha-tropiques-technical-memorandum/> (last access: 5 November 2019), 2011.
- Tabata, T., John, V. O., Roebeling, R. A., Hewison, T., and Schulz, J.: Recalibration of over 35 Years of Infrared and Water Vapor Channel Radiances of the JMA Geostationary Satellites, *Remote Sens.*, 11, 1189, <https://doi.org/10.3390/rs11101189>, 2019.
- Wolff, T.: An image geometry model for ME-TEOSAT, *Int. J. Remote Sens.*, 6, 1599–1606, <https://doi.org/10.1080/01431168508948308>, 1985.