



TOR VERGATA UNIVERSITY, ROME, ITALY

Department of Civil Engineering and Computer Science Engineering

Computer Science, Control and Geoinformation PhD Programme  
XXVIII Cycle

**Potential of EUMETSAT MTG-IRS  
hyperspectral sounder for improving nowcasting  
and very short range forecast atmospheric models**

**Antonio Vocino**

Advisor: **Prof. Fabio Del Frate**  
Coordinator: **Prof. Giovanni Schiavon**

A dissertation submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy

Academic Year 2018/19



*Antonio Vocino*

Potential of EUMETSAT MTG-IRS  
hyperspectral sounder for improving nowcasting  
and very short range forecast atmospheric models

DOCTORAL THESIS IN COMPUTER SCIENCE,  
CONTROL AND GEOINFORMATION

©2019 Antonio Vocino - antonio.vocino@gmail.com

available online at:

*[http://www DISP.unroma2.it/geoinformation/students/geoinformation-dissertations/Thesis\\_Vocino.pdf](http://www DISP.unroma2.it/geoinformation/students/geoinformation-dissertations/Thesis_Vocino.pdf)*

This work is licensed under a Creative Commons Attribution 3.0 Italy License.

To view a copy of this license visit: *<http://creativecommons.org/licenses/by/3.0/it>*



*To my wife Serena  
and my children Marco Elena and Davide.*

*You are the sunshine of my life.*



# ACKNOWLEDGEMENTS

---

Approaching the end of this work I would express my gratitude to the many people who have been important to me in this challenge.

First of all, I am grateful to my PhD advisor, Professor Fabio Del Frate, for supporting me since the beginning, for sharing and enlightening ideas and tips on my research topic and for trusting me all this time. I am thankful to Professor Giovanni Schiavon as well, for his supervision and coordination of activities during the doctorate.

In the same beautiful frame of Rome Tor Vergata University, I would thank all the Earth Observation Laboratory people, who take care of sharing the knowledge on such a complex but amazing field as Geoinformation is. Someone arrives, someone leaves, but science stays in that room.

I am particularly grateful to the IRS EUMETSAT expert teams I was involved with for many years, namely the MIST before and the MIMAG after, for the valuable and outstanding skills and the friendly atmosphere that gave me so much. Many hints for the research performed in this thesis come from there. Rolf and Stephen, I will remember your words forever.

This work would not have been possible without the support and facilities provided by the Italian Air Force. The Meteorological Service gave me the opportunity to dedicate part of my time – at work and at university – to such topical matter and to be in touch with the real core of the IRS mission development. At the Meteorological Centre in Pratica di Mare I can't forget the wise teachings and the absolute trust of Lucio Torrisi.

A special thanks I would express to my twin of the goal, Daniele Biron, for believing in me as nobody else from the very beginning of the story, for his continuous encouragement, infinite source of knowledge, patience, sense of irony and laughs.

Finally, I am forever thankful to my wife Serena for all of the love and support, even when I have brought home work and I was mentally distracted or my working commitments have kept me away. Thanks so much to my wonderful children Marco Elena and Davide for being exactly as they are, for drawing a smile on my face every single day. Vi amo con tutto me stesso.

*Antonio Vocino*

*Rome, March 2019*





*Creativity is just connecting things.*

*Steve Jobs*



---

## Potential of EUMETSAT MTG-IRS hyperspectral sounder for improving nowcasting and very short range forecast atmospheric models

### **Short abstract:**

The main objective of this thesis is to investigate the potential of the future EUMETSAT Meteosat Third Generation - InfraRed hyperspectral Sounder (MTG-IRS) products for improving nowcasting and Very Short Range Forecast (VSRF) systems, focusing in particular on the operational applications in modern National Meteorological Services (NMSs). The thesis is mainly founded on the work carried out by the author in the framework of MTG-IRS Near Real Time Demonstration Project, conceived and led by EUMETSAT to enhance the user awareness on the potential of the IRS instrument in support to the meteorology and in particular to the nowcasting activities.

**Keywords:** meteorology, remote sensing, IR hyperspectral sounding, numerical weather prediction, nowcasting, very short range forecast, geoinformation, spatial decision support systems

---



# CONTENTS

---

<b>Abstract</b>	<b>v</b>
<b>Preface</b>	<b>vii</b>
<b>List of Abbreviations</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	2
1.2 Motivation . . . . .	2
1.3 Scope of the work . . . . .	5
1.4 Geoinformation contribution . . . . .	5
<b>2 The IRS instrument</b>	<b>7</b>
2.1 Meteosat Third Generation Programme . . . . .	8
2.2 MTG observation missions and payloads . . . . .	11
2.3 IRS instrument . . . . .	14
2.3.1 Description . . . . .	14
2.3.2 User requirements and expected benefits . . . . .	18
<b>3 Concepts and methods</b>	<b>21</b>
3.1 The potential of IR hyperspectral sounding . . . . .	22
3.2 IASI interferometer as proxy for IRS . . . . .	26
3.3 The recipe of this project . . . . .	29
<b>4 The MTG-IRS proxy data Italian Demonstration Project</b>	<b>33</b>
4.1 EUMETSAT IRS NRT demonstration project . . . . .	34
4.2 ItDP software package . . . . .	36
4.3 An example of IASI data processing . . . . .	38
4.4 Examples of IRS proxy data processing . . . . .	42
<b>5 Project Methodology</b>	<b>45</b>
5.1 Software Architecture . . . . .	46
5.2 Functionalities . . . . .	51
5.2.1 Processing System . . . . .	51
5.2.2 Supervisor System . . . . .	56
5.2.3 External Libraries . . . . .	57
5.3 An example of end-to-end processor . . . . .	58
5.4 What's next? . . . . .	60
<b>6 Case studies</b>	<b>63</b>
6.1 Colocation and analysis of the problem . . . . .	65

---

6.2	Revisiting time . . . . .	70
6.3	Object detection . . . . .	73
6.4	Impact of a priori state . . . . .	76
<b>7</b>	<b>Discussion and conclusions</b>	<b>81</b>
7.1	Considerations . . . . .	82
7.2	Applications . . . . .	84
7.3	Further developments . . . . .	85
7.4	Concluding remarks . . . . .	86
	<b>Bibliography</b>	<b>87</b>
	<b>List of Figures</b>	<b>89</b>
	<b>List of Tables</b>	<b>93</b>

---

**Abstract:**

In this thesis the research activities aiming at the investigation on the use of hyperspectral IR data for the diagnosis of atmospheric instability and the early detection of convective systems are shown.

The study was carried out in the framework of MTG-IRS Near Real Time Demonstration Project, conceived and led by EUMETSAT to enhance the user awareness on the potential of the IRS instrument in support to the meteorology and in particular to the nowcasting activities.

In detail, the proxy IRS hyperspectral level 2 products, generated from real IASI and CrIS data and distributed by EUMETSAT, were processed in near real time together with auxiliary colocated and independent datasets to assess the correlation between the signal (i.e. the information content of level 2 products) and the weather phenomenon (convective instability).

The reprocess of a set of significant case studies over Italy was also included in the study. Research results show that the exploitation of hyperspectral data in the field of nowcasting applications could enhance the capacity and user-readiness of modern, operational Meteorological Services with respect to the early detection of severe weather.

---

---

## **Riassunto:**

Obiettivo delle attività di ricerca descritte in questa tesi è lo studio dell'utilizzo dei dati iperspettrali IR per la diagnosi dell'instabilità atmosferica ed il rilevamento anticipato di sistemi convettivi.

Lo studio è stato condotto nell'ambito del progetto MTG-IRS Near Real Time, concepito e coordinato da EUMETSAT per potenziare la preparazione degli utenti sulle potenzialità dello strumento IRS a supporto della meteorologia ed in particolare delle attività di previsioni a brevissima scadenza.

In dettaglio, i prodotti iperspettrali di livello 2 di IRS, generati a partire da dati reali di IASI e CrIS e distribuiti da EUMETSAT, sono stati processati in quasi tempo reale insieme a dati ausiliari geograficamente co-localizzati ed indipendenti al fine di valutare la correlazione tra il segnale (cioè il contenuto informativo dei prodotti di livello 2) ed il fenomeno meteorologico (l'instabilità convettiva).

Lo studio comprende anche il riprocessamento di una serie di casi di studio significativi sull'Italia. I risultati della ricerca mostrano che lo sfruttamento dei dati iperspettrali nel settore delle previsioni a brevissima scadenza è in grado di potenziare la capacità e la prontezza a livello utente dei moderni Servizi Meteorologici operativi per quanto riguarda il rilevamento in anticipo dei fenomeni intensi.

---



# PREFACE

---

This thesis was submitted to the Department of Civil Engineering and Computer Science of the University of Rome Tor Vergata as a partial fulfillment of the requirements to obtain the PhD degree. The work presented was carried out in the years 2012-2018 as research activity without scholarship for the Geoinformation Doctorate under the mentorship of Prof. Fabio Del Frate.

In this regard, I've always enjoyed my experience as part-time student at the Tor Vergata University Geoinformation Doctoral School, trying to assimilate the state-of-art Knowledge (with capital K) in this huge multi-disciplinary field. Geoinformation encompasses indeed remote sensing, imaging, positioning, geospatial analysis, geomatics, electromagnetics and Earth sciences. My ambition was therefore to bring new concepts and ideas back to my job, namely as meteorologist and atmospheric physicist for the Italian Air Force Meteorological Service.

Under such auspices, the thesis should be viewed as a *repository* of new areas of investigation aiming at the application of geoinformation pieces of knowledge to the fundamental problem here addressed, concerning the best exploitation of the information content of the future IRS hyperspectral sounder, expected in orbit in 2022.

## THESIS CONCEPT AND OBJECTIVES

The original concept of the research project reported in this thesis is based on my work carried out as member of the EUMETSAT MTG-IRS Mission Advisory Group (MIMAG). As proof of concept, the experimental datasets made available as proxy data generated in near real-time (NRT) by EUMETSAT from true radiance data measured by IASI sounder on board METOP A/B satellites and by CrIS on board Suomi NPP have been processed and analysed.

The main objectives of the investigation are primarily to 1) study the IRS instrument, to characterise the information content of both level 1 (radiance

data) and level 2 (geophysical variables, namely retrieved vertical profiles of temperature and humidity) and consequently to set up the optimal processing aiming at filtering out false alarms and at detecting the atmospheric instability signal, 2) explore innovative concepts and methods (e.g. NN, cluster analysis) largely used and developed within expert systems, but not yet sufficiently evaluated for operational use in the framework of very short-term weather prediction (*nowcasting*, more properly) and 3) design and set up a modular software prototype to process the proxy data, configured and run on the computing resources of the Italian Air Force Meteorological Centre, able to elaborate level 2 and ancillary data in NRT into an operational context, as technology demonstrator to introduce its output products in modern geoinformation systems for decision support.

## THESIS OUTLINE

This thesis is structured as follows. Chapter 1 introduces background knowledge and highlights the motivation of the research conducted in the thesis. Chapter 2 provides a description of the future MTG-IRS, focusing in particular on the expected benefits in terms of capability of detection of the atmospheric instability. Chapter 3 describes the scientific environment and the preliminary experiments on the use of IR hyperspectral soundings for nowcasting applications, introducing the theoretical concepts and the experimental methods used throughout the thesis. Chapter 4 provides a description of the demonstration project conceived and realised by myself in support of the EUMETSAT MTG-IRS Mission Advisory Group activities. Chapter 5 provides the exact details on the original procedures designed, implemented and tested to achieve the aim of the project, as baseline methodology for ItDP and future developments. Chapter 6 provides an overview of the significant case studies collected within ItDP project and selected to highlight some key features of IRS proxy data and their potential impact on nowcasting applications. Finally, chapter 7 summarises the findings of the investigation conducted in the thesis and provides the concluding remarks, giving an outlook on possible applications and future work.

# LIST OF ABBREVIATIONS

---

<b>AIRS</b>	Atmospheric Infrared Sounder
<b>ATBD</b>	Algorithm Theoretical Basis Document
<b>BUFR</b>	Binary Universal Form for the Representation of meteorological data
<b>CAF</b>	Central Application Facility
<b>CAPE</b>	Convective Available Potential Energy
<b>CIMSS</b>	Cooperative Institute for Meteorological Satellite Studies
<b>CNMCA</b>	Centro Nazionale di Meteorologia e Climatologia Aeronautica
<b>COMET</b>	Centro Operativo per la Meteorologia
<b>COSMO</b>	COnsortium for Small-scale MOdeling
<b>CrIS</b>	Cross-track Infrared Sounder
<b>DBSCAN</b>	Density-Based Spatial Clustering of Applications with Noise
<b>ECMWF</b>	European Centre for Medium range Weather Forecasts
<b>EDA</b>	Ensemble Data Assimilation
<b>EPS</b>	EUMETSAT Polar System
<b>ESA</b>	European Space Agency
<b>EUMETSAT</b>	European Organisation for the Exploitation of Meteorological Satellites
<b>EURD</b>	End-User Requirements Document
<b>FCI</b>	Flexible Combined Imager
<b>FOV</b>	Field of View
<b>FTS</b>	Fourier Transform Spectrometer
<b>GEO</b>	GEOstationary orbit
<b>GIFTS</b>	Geosynchronous Imaging Fourier Transform Spectrometer
<b>GII</b>	Global Instability Index
<b>GIS</b>	Geographic Information System
<b>GMES</b>	Global Monitoring for Environment and Security
<b>GOES</b>	Geostationary Operational Environmental Satellite
<b>GOS</b>	Global Observing System
<b>GRIB</b>	General Regularly-distributed Information in Binary form
<b>HDF</b>	Hierarchical Data Format
<b>HPC</b>	High Performance Computing
<b>IASI</b>	Infrared Atmospheric Sounding Interferometer
<b>ICT</b>	Information and Communication Technology
<b>IDPF</b>	Instrument Data Processing Facility

<b>IRS</b>	InfraRed Sounder
<b>ItDP</b>	Italian Demonstration Project
<b>JSON</b>	JavaScript Object Notation
<b>L1</b>	Level 1
<b>L2</b>	Level 2
<b>L2PF</b>	Level 2 Processing Facility
<b>L2VDP</b>	Level 2 Validation and Demonstration Processor
<b>LAC</b>	Local Area Coverage
<b>LAM</b>	Limited Area Model
<b>LEO</b>	Low Earth Orbit
<b>LI</b>	Lightning Imager
<b>LWIR</b>	Long Wave InfraRed radiation
<b>MCS</b>	Mesoscale Convective System
<b>METOP</b>	Meteorological Operational satellite
<b>MIMAG</b>	Meteosat third generation Infrared sounder Mission Advisory Group
<b>MIST</b>	Meteosat third generation Infrared sounder Science Team
<b>ML</b>	Machine Learning
<b>MODE</b>	Method for Object-based Diagnostic Evaluation
<b>MOPD</b>	Maximum Optical Path Difference
<b>MRD</b>	Mission Requirements Document
<b>MS</b>	Meteorological Service
<b>MSG</b>	Meteosat Second Generation
<b>MTG</b>	Meteosat Third Generation
<b>MTG-I</b>	MTG-Imaging satellite
<b>MTG-S</b>	MTG-Sounding satellite
<b>MWIR</b>	Mid Wave InfraRed radiation
<b>NASA</b>	National Aeronautics and Space Administration
<b>NetCDF</b>	Network Common Data Form
<b>NIR</b>	Near InfraRed radiation
<b>NN</b>	Neural Network
<b>NOAA</b>	National Oceanic and Atmospheric Administration
<b>NRT</b>	Near Real Time see [CONV] for a definition
<b>NWC</b>	Nowcasting
<b>NWP</b>	Numerical Weather Prediction
<b>PCA</b>	Principal Component Analysis
<b>RSS</b>	Rapid Scanning Service
<b>RTM</b>	Radiative Transfer Model

<b>SAF</b>	Satellite Application Facility
<b>SDLC</b>	Systems Development Life Cycle
<b>SDSS</b>	Spatial Decision Support System
<b>SEVIRI</b>	Spinning Enhanced Visible and InfraRed Imager
<b>SRD</b>	System Requirements Document
<b>Suomi-NPP</b>	Suomi National Polar-orbiting Partnership satellite
<b>SWIR</b>	Short Wave InfraRed radiation
<b>UKMO</b>	United Kingdom Meteorological Office
<b>UV</b>	UltraViolet radiation
<b>UVN</b>	Ultraviolet, Visible & Near-infrared sounder
<b>VIS</b>	VISible radiation
<b>VSRF</b>	Very Short Range Forecasting
<b>WMO</b>	World Meteorological Organization



# 1

## INTRODUCTION

---

*This chapter provides background knowledge and highlights the motivation of the research conducted in the thesis.*

### Contents

---

1.1	Background . . . . .	2
1.2	Motivation . . . . .	2
1.3	Scope of the work . . . . .	5
1.4	Geoinformation contribution . . . . .	5

---

## 1.1 Background

---

The current satellites of the Meteosat Second Generation (MSG) system have been delivering reliable and accurate observations and early warning services since 2002, in support of the detection and forecast of rapidly developing high impact weather, such as thunderstorms or fog. Nowcasting and very short range forecasts — up to twelve hours — are vital for the safety of life, property and infrastructure and rely on very frequent, detailed images of the atmosphere.

EUMETSAT next generation of geostationary meteorological satellites, Meteosat Third Generation (MTG) will continue services beyond this decade and address future challenges in weather forecasting and other services for European citizens, such as improved air quality or UV-radiation warnings, as well as climate and atmospheric chemistry monitoring. With the planned launch of six new geostationary 3-axis-stabilised satellites in twin configuration (for imaging and sounding missions) from 2021 onwards, MTG will provide an evolution of the imaging service, including a new Lightning Imager, and a state-of-the-art atmospheric sounding service providing measurements in the infrared and ultraviolet spectrum.

MTG will carry within the sounding payload an InfraRed hyperspectral Sounder (IRS). The IRS mission is intended to provide in real time, for the first time ever, a four-dimensional (over time and space) high-resolution data on water vapour and temperature structures.

Thanks to the scan strategy and timeliness of the instrument and its spectral resolution, IRS level 1 (spectral) and level 2 (geophysical) data will deliver unprecedented information on horizontal and vertical gradients of moisture, wind and temperature – in particular in pre-convective tropospheric environment – which play a crucial role for the initiation of severe weather phenomena. Therefore, from the Numerical Weather Prediction (NWP) perspective, the exploitation of the IRS information content is expected to be extremely beneficial [1].

## 1.2 Motivation

---

Nowcasting and very short range weather forecast applications are largely based on observational data, to provide guidance to meteorologists making



detailed information on the 3-Dimensional atmospheric state [2]. Sophisticated nowcasting techniques are now routinely used in developed countries where complex observing systems as meteorological radar are mature and robust. However, in less developed countries the required operational radar systems needed for nowcasting are still missing and satellite observations are often the only source of information available for monitoring the weather.

Contrary to a Numerical Weather Prediction (NWP) data assimilation system, where the goal is to estimate each atmospheric variable on a more or less regular grid, synoptic meteorologists and nowcasters attempt to depict meteorological phenomena in an “object-oriented” way. Forecasting methods evolve from looking at individual observing systems separately to integrating different data sources in order to infer meteorological parameters and phenomena, based on the preliminary evaluation of the impact of the different observing systems.

Geostationary satellites imagery is the prime source for locating synoptic-scale features and objects in real-time, allowing them to detect any incipient discrepancies between model forecasts and reality at an early stage. This is particularly true over oceanic areas, where conventional data are typically very sparse. Model fields and satellite imagery may be superposed on a workstation screen; a good example is given by the potential vorticity field of the upper-level flow correlated with water vapour satellite images. The horizontal resolution and coverage are good, except over the Polar Regions (60-90N and 60-90S). The vertical resolution has been enhanced with the new generation of geostationary satellites GOES and Meteosat.

The steadily improving horizontal and spectral resolution of geostationary satellite instruments leads to improved detection and classification of clouds. Progress has been made on night-time detection of low clouds, which used to be marginal, and on distinction between high-thin and high-thick clouds (e.g. cirrus versus cumulonimbus). Many of the derived products are very useful for nowcasting purposes, and the enhanced rapid-scan facility on Meteosat is extremely beneficial. Furtherly, nowadays quantitative precipitation estimates from geostationary satellites are almost comparable with the ones provided by surface-based weather radar networks.

Meteorological satellite data are well suited to monitoring in a qualitative way the initiation and rapid development of precipitation generating systems both in space and time. Rapid imaging (on the order of minutes) is critical to nowcasting, even though this service competes with broader coverage requirements. Frequent images from geostationary satellites provide good to adequate horizontal resolution for identifying the initiation, evolution and movement of synoptic and mesoscale cloud systems as well as of local circulations over most of the tropics and temperate zones.

The more frequent and more comprehensive data collected by latest satellites will also aid the weather forecasters in the fast recognition and accurate prediction of dangerous weather phenomena such as heavy thunderstorms, thus forming an important contribution to nowcasting. Indeed, the air mass parameters (moisture, temperature at different tropospheric levels) derived from satellite are currently used to issue severe weather warnings, if a derived severe weather index exceeds a certain threshold. These thresholds are usually determined empirically and should not be regarded as fixed values. A skilled local forecaster is still absolutely necessary for a correct interpretation, but of course an operational satellite-based retrieval of these crucial parameters can provide a great potential for the identification of pre-convective conditions.

In particular, the class of instruments operating in the InfraRed (IR) region of the spectrum and known as IR hyperspectral sounders, currently on board polar-orbiting satellites, is considered nowadays one of the most gap-filling observing system for weather forecast in general.

As of today, data from IR hyperspectral sounders (IASI on METOP, CrIS on Suomi-NPP) are widely used, at the operational meteorological centres, to characterise air masses and clouds and to monitor the atmospheric instability. However, IASI/CrIS observations are not used massively for operational nowcasting systems and decision-aid applications. The main limiting factors, due to the polar orbit, are the small number of overpasses per day and the timeliness issues (related to the so called cut-off time), disagreeing with the basic NWC requirements to detect and track the small scale (in space and time) moisture changes.

Conversely, the MTG-IRS 30 minutes repetition cycle, allowed by the geostationary orbit, is strategic for NWC and VSRF applications, especially for monitoring the pre-convective weather parameters such as stability indices, tendencies, moisture convergence. Moreover, the IRS high resolution level 1 spectra reveal absorption lines providing directly information on low level moisture gradients, low level inversions, tropopause inversions, which are key factors for triggering severe storms.

Finally, the derived level 2 products provide information on vertical and horizontal moisture and temperature structures, surface properties (temperature, emissivity), and cloud properties (altitude, temperature, optical thickness, ice/liquid content), which are easily assimilated into NWP operational systems.

## 1.3 Scope of the work

---

The study for propaedeutic activities related to the investigation of the potential of the MTG-IRS products for improving Nowcasting (NWC) and Very Short Range Forecast (VSRF) atmospheric models is the main objective of this work, focusing in particular on the operational applications in modern Meteorological Services (MSs). Moreover, a by-product of the research analysis is the possibility of exploring innovative concepts and methods (e.g. NN, cluster analysis) largely used and developed within expert systems, but not yet sufficiently evaluated for operational use in the field of the atmospheric nowcasting, for the exploitation of the current and future hyperspectral data in support of the early detection of severe weather systems.

## 1.4 Geoinformation contribution

---

This study is developed in the research framework of the University of Rome-Tor Vergata - Doctoral School in Computer Science, Control and Geoinformation where the presenting author is undertaking its PhD. Thanks to this opportunity, a huge contribution in terms of ideas and state-of-art knowledge from this world is gathered into the research work, with particular emphasis on the Machine Learning (ML) theory and applications and on the programming skills. This brought the author to design and set up a modular software prototype to process and analyse all the heterogeneous datasets involved in the project, as for example level 2 and ancillary data elaborated in an operational context, paving the way for a technology demonstrator to translate the results of this work in modern geoinformation systems for decision support.



# 2

## THE IRS INSTRUMENT

---

*This chapter provides a description of the future MTG-IRS, focusing in particular on the expected benefits in terms of capability of detection of the atmospheric instability.*

### Contents

---

<b>2.1</b>	<b>Meteosat Third Generation Programme . . . . .</b>	<b>8</b>
<b>2.2</b>	<b>MTG observation missions and payloads . . . . .</b>	<b>11</b>
<b>2.3</b>	<b>IRS instrument . . . . .</b>	<b>14</b>
2.3.1	Description . . . . .	14
2.3.2	User requirements and expected benefits . . . . .	18

---

---

## 2.1 Meteosat Third Generation Programme

---

Meteosat Third Generation (MTG) is the next-generation European operational geostationary meteorological satellite system – a collaborative EUMETSAT/ESA programme to continue and enhance the services offered by the Second Generation of the Meteosat System (MSG). Preparatory activities for the EUMETSAT MTG series started in late 2000 in cooperation with ESA, following the decision of the EUMETSAT Council to proceed with a Post-MSG<sup>1</sup> User Consultation Process. The process was aiming at capturing the foreseeable needs of users of EUMETSAT’s satellite data in the timeframe of 2015-2035 [3].

By the way, typical development cycles of complex space systems are on the order of a decade or more. The current MSG system is expected to deliver observations and services until 2024. MTG will continue services beyond this date<sup>2</sup> and address future challenges in weather forecasting and other services for European citizens, such as improved air quality or UV-radiation warnings, as well as climate and atmospheric chemistry monitoring.

The early consultation process was implemented through the setup of Applications Expert Groups (AEGs) tasked to propose their vision on operational services and information systems in the cited timeframe and to define associated needs and priorities for input information and observations. The AEGs concentrated on two families of applications, defined in sufficiently broad terms to address even non-meteorological applications:

### **Nowcasting and Very Short Range Forecasting (VSRF)**

defined in a very broad sense as “user-driven services using appropriate meteorological and related science to provide information on expected conditions up to 12 hours ahead” and hence covering e.g. air pollution, ocean, hydrology at these timescales;

### **Numerical Weather Prediction (NWP)**

including marine forecasting and emerging applications like extended range weather forecasting and seasonal prediction.

This last step of the process produced the inputs required to support the MTG definition studies (pre-phase A) jointly planned by ESA and EUMETSAT for 2004, synthesised in the MTG Mission Requirements Document (MRD) [4].

---

<sup>1</sup>The initial name of the MTG programme.

<sup>2</sup>As of today, the first MTG I/S spacecrafts are planned to be in orbit in 2021/23.

For the space segment, the early definition studies and consultations identified the following five observation missions for MTG:

- Full Disc High Spectral resolution Imagery (FDHSI);
- High Resolution and Fast Imagery (HRFI);
- Lightning detection Mission (LI);
- IR Sounding Mission (IRS);
- UV-VIS-NIR Sounding Mission as a payload complement (UVN).

After pre-phase A mission studies, where preliminary instrument concepts were investigated allowing the consolidation of the most critical and demanding technical requirements, phase A studies (2006-2008) addressed both the space segment system feasibility and ground and operations programmatic aspects [5].

Regarding the space segment, in particular, the phase A investigation covered the entire suite of optical instruments identified in the preliminary assessments including:

- 1) a flexible combined imager for both FDHSI and HRFI missions;
- 2) a Fourier Transform Spectrometer for IRS observations;
- 3) a lightning detector sensor.

For the sake of completeness, the study of concepts and implementation of the UVN mission of MTG are covered by the efforts of ESA and the European Union (EU) in the framework of the Global Monitoring for Environment and Security (GMES) Sentinel 4 program.

In October 2008, an agreement was signed between ESA and EUMETSAT agencies regulating the respective roles, responsibilities and financial commitments of the two organizations concerning future phases of the MTG programme as well as the approval of the MTG payload complement:

- EUMETSAT will specify and consolidate the end-user requirements, the overall mission requirements, the space-to-ground requirements, and the ground segment requirements. In addition, EUMETSAT will be responsible for the overall mission and system engineering and ground segment design and development. Further along, it will fund the procurement of the recurring satellites, the launch services and launch and early orbit phases, and also execute commissioning and operations.

- ESA is responsible for the development and implementation of the space segment technologies and the first MTG twin satellites. ESA is funding the related cost, apart from 30% to be contributed by EUMETSAT. ESA will procure all recurring satellites as part of the EUMETSAT development and operations programme.

In March 2010, the MTG program took another step towards full approval with the EUMETSAT special Council, by accepting the MTG End-User Requirements Document (EURD) [6] which defines not only the deliverables to the user community, but also the duration of the operational service (at least 20 years for the imagery mission and at least 15.5 years for the sounding mission), the number of satellites and their in-orbit lifetime.

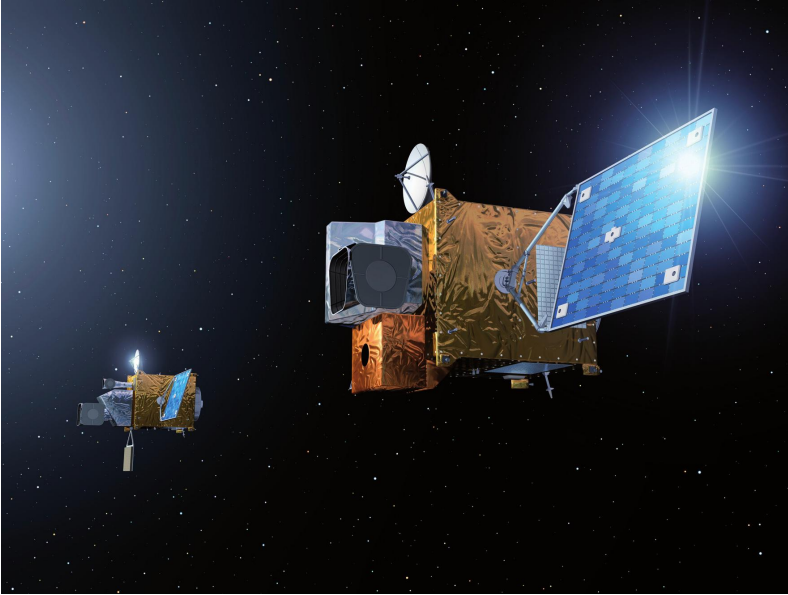
In 2011 the full MTG programme entered into force, after the System Requirements Review (SRR) for the space segment followed by the Preliminary Design Review (PDR) which closed the Phase B activities and led to the current Phase C/D of detailed definition, development, production and testing of all MTG system elements. The timeline of the phases of MTG Programme is summarised in table 2.1.

**Table 2.1:** MTG overall timeline

Timeframe	Activities
2001-2005	User Consultation Process & Pre-Phase-A Studies (Phase 0): - 2001-2003: Phase 1 - High-level user needs & priorities agreed, preparation of Pre-Phase-A studies; - 2004-2005: Phase 2 - system concept studies (Pre-Phase-A), evaluation/pre-selection of MTG missions.
2006-2008	MTG (Phase-A): Feasibility studies of selected mission concepts. Approval process of MTG preparatory program. The Phase-A studies were successfully concluded in December 2008.
2009-2010	Phase-B: Detailed design activities under coordinated EUMETSAT and ESA Preparatory Programs, with approval processes for coordinated MTG development programs.
2011-2022	Phases C/D: Development and testing of the MTG system elements.
2022 onwards	Phase E: Nominal need date for MTG series. Launch, operations and utilization of MTG spacecraft.



## 2.2 MTG observation missions and payloads

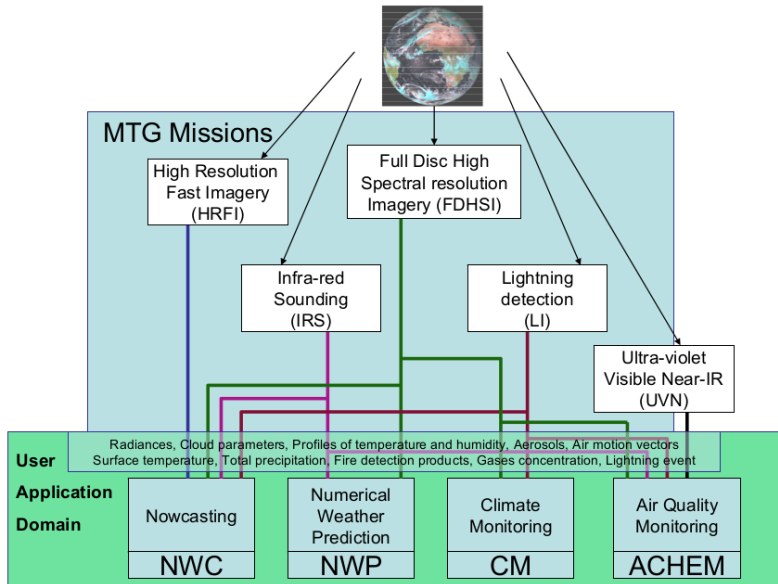


**Figure 2.1:** Artist's rendition of the MTG-S (in the foreground) and MTG-I series spacecraft in orbit (image credit: ESA).

According to MTG-EURD [6], to fulfil MTG mission it is required to deploy sustained capabilities to acquire, process and distribute to downstream users and processing centres observations and geophysical parameters of the Earth/atmosphere system on a broad spectral range (from UV to LWIR), covering extensive areas (global and regional) at high spatial, spectral and temporal resolution.

To accomplish this, MTG mission encompasses the following observation missions (figures 2.2 and 2.3):

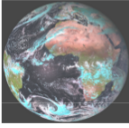

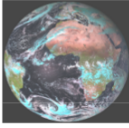
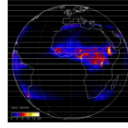
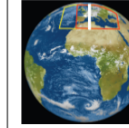
- Flexible Combined Imager (FCI) mission, allowing to scan either the full disc in 16 channels every 10 minutes with a spatial sampling distance in the range 1 – 2 km (Full Disc High Spectral resolution Imagery (FDHSI) in support of the Full Disc Scanning Service (FCI-FDSS)) or a quarter of the earth in 4 channels every 2.5 minutes with a resolution twice better (High spatial Resolution Fast Imagery (HRFI) in support of the Rapid Scanning Service (FCI-RSS));



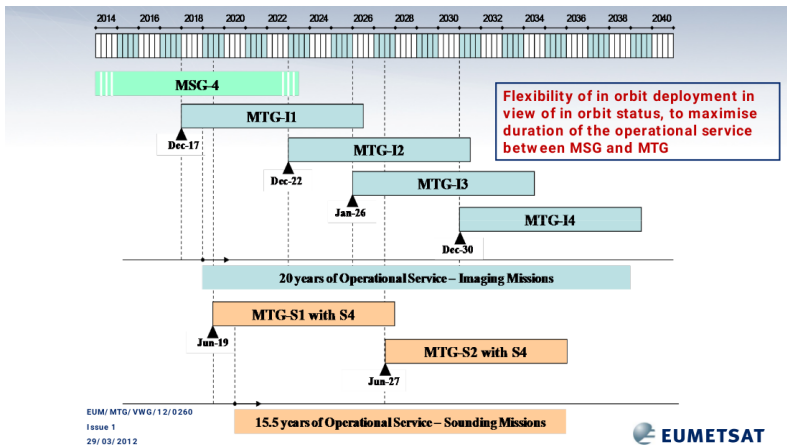
**Figure 2.2:** MTG missions and main user applications beneficiaries (from [5]).

- InfraRed Sounder (IRS) mission, covering the full disc in 60 minutes, providing hyperspectral sounding information in two bands, Long Wave InfraRed (LWIR) and Mid Wave InfraRed (MWIR), with a spatial sampling distance around 4 km (see next section);
- Lightning Imager (LI) mission, detecting continuously, over almost the full disc, the lightning discharges taking place in clouds or between cloud and ground with a spatial sampling distance around 10km;
- Ultraviolet, Visible & Near-infrared (UVN) sounding mission, covering Europe every hour taking measurements in three spectral bands (UV: 290 – 400 nm; VIS: 400 – 500 nm, NIR: 755 – 775 nm) with a spatial sampling distance around 10 km. The UVN mission is implemented with the GMES Sentinel-4 instrument accommodated in the MTG-S satellites.

The Space Segment of the MTG System consists of a satellites constellation. Three in-orbit satellites are needed to support the complete set of missions and functions listed above, the so-called Full Operational Capability (FOC). To span the operational life time of the programme over at least 20 years, MTG will see the launch of six new geostationary (imaging and sounding) satellites from 2021 onwards (see figure 2.4).

				
Full Disc High Spectral resolution Imagery	High spatial Resolution Fast repeat Imagery	Infra-red Sounding hyperspectral Imagery	Lightning detection	Ultra-violet Visible and Near-IR mission
10 minutes	2.5 minutes	60 minutes	continuous	60 minutes

**Figure 2.3:** MTG missions spatial coverage and temporal rate (from [5]).



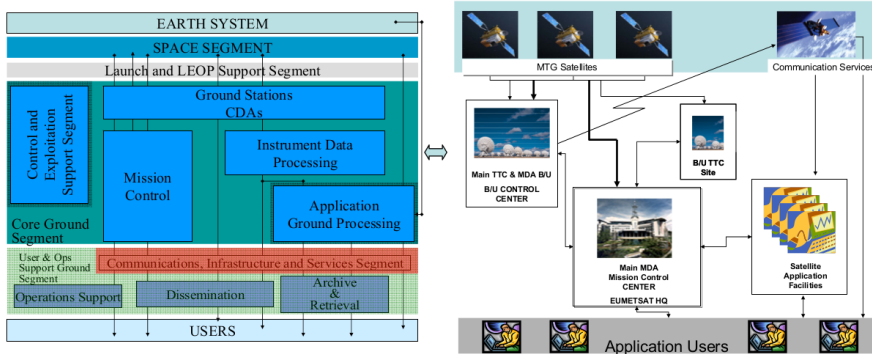
**Figure 2.4:** MTG in orbit deployment scenario. A temporal shift of the whole programme is planned, thanks to the extension of the operational life of MSG satellites<sup>2</sup> (image credit: EUMETSAT).

The satellite series will be based on 3-axis platforms and comprise:

- 4 Imaging Satellites (MTG-I) (20 years of operational services expected);
- 2 Sounding Satellites (MTG-S) (15.5 years of operations expected).

The three in-orbit satellites will deliver the full disc service (prime MTG-I, nominally at  $0^\circ$  longitude), the Rapid Scanning Service (second MTG-I satellite, expected to be at  $9.5^\circ E$  longitude and acting also as in-orbit hot backup for the prime MTG-I) and the new sounding services by the MTG-S satellite.

The overall picture of the MTG system segments, including the set of the involved operational centres is shown in figure 2.5.



**Figure 2.5:** MTG system segments and operational sites (from [5]). The overall system is divided in five segments: Space, Launch and Early Operation, Core Ground Segment, Application Ground Segment, User and Operations Support (left). The system components will be distributed over various sites to pursue the highest resilience. Commercial communication satellite services will be used, as currently in the EUMETCAST system, for dissemination of data to regional and global users (right).

## 2.3 IRS instrument

### 2.3.1 Description

The MTG-IRS (Meteosat Third Generation InfraRed Sounder) is a Michelson interferometer infrared sounder, belonging to the class of so-called Fourier Transform Spectrometers (FTS), measuring the total Earth-atmosphere system upwelling thermal radiation. Further details on the theoretical aspects and the applications of FTS in the remote sensing field can be found in [7] and [8]. According to the EUMETSAT technical specifications [1], IRS shall use a Maximum Optical Path Difference (MOPD) of at least 0.8 cm, covering the IR spectral range in two spectral bands, namely the LWIR band (spectral domain  $700 - 1210 \text{ cm}^{-1}$ ) and the MWIR band (spectral domain  $1600 - 2175 \text{ cm}^{-1}$ ). These two bands, shown in figure 2.7, are sampled at moderately high spectral rate of  $0.625 \text{ cm}^{-1}$  (1738 channels in total)<sup>3</sup> [9].

<sup>3</sup>It is worth to note that the IRS instrument will provide observations over an extended spectral domain, namely from  $680 - 1210 \text{ cm}^{-1}$  and from  $1600 - 2250 \text{ cm}^{-1}$ . However for the spectral channels outside of the nominal domain defined in the text above, no radiometric performance requirements have been specified, thus it is not clear how the

The MTG-IRS instrument has no direct MSG heritage but is closely related, from the user point of view, to the operational EPS IASI mission on board METOP series satellites and, from a more general point of view, to the NASA GIFTS mission conceived in the USA in the framework of the third millennium programme. The breakthrough characteristics of IRS concern, of course, its high horizontal resolution (about 4 km) coupled with high vertical and spectral resolution. Regarding the fourth (temporal) dimension the geostationary orbit provides, together with the scanning strategy, an unprecedented revisiting time, performing observations of a quarter of the Earth disc in just 15 minutes.

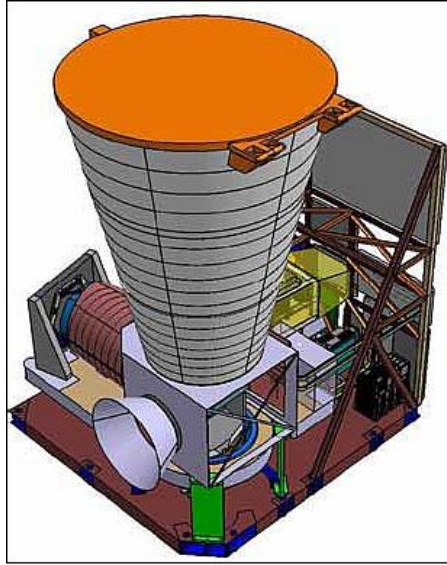
The instrument scans the geo disk in a “step and stare” mode. During each stare a portion of the scene, called “dwell”, is observed by a two dimensional array of  $160 \times 160$  detectors measuring the upwelling terrestrial radiation for about 10 seconds, resulting in 2560 spectra per second acquired simultaneously. The dwell coverage is stepped in an east/west direction to form a line of dwell spectral soundings, before moving northward to form the next line, covering the Local Area Coverage (LAC) within the repeat cycle duration. The full geo disc is divided into four LAC zones. Each local area scanning zone represents an area which can be scanned in 15 minutes. An illustration of these four scanning zones is given in figure 2.8, showing that to complete the full disc scan a basic repeat cycle of 6-hours is required. However, as this scan scenario has been designed to provide frequent observations over Europe, LAC-4 (shown in figure 2.9) is observed every 30 minutes within a 15 minutes period. As of today, the operational scanning sequence<sup>4</sup>, repeated every 6 hours, has been approved according to the following patterns: 3 times (LAC3 + LAC4) followed by 3 times (LAC2 + LAC4) followed by 3 times again (LAC3 + LAC4) and ultimately 3 times (LAC1 + LAC4) as shown in figure 2.10.

Before the observations are transmitted to ground, a number of operations are performed. The measurements are quality controlled and tested for errors, in particular spikes in the interferograms. They are also corrected for deformations of the interferograms, induced by the off-axis position of the detector elements and for non-linear response of the detectors themselves. The spectral soundings are then transmitted as decimated interferograms to the ground segment, including the receiving stations which send them to the Instrument Data

---

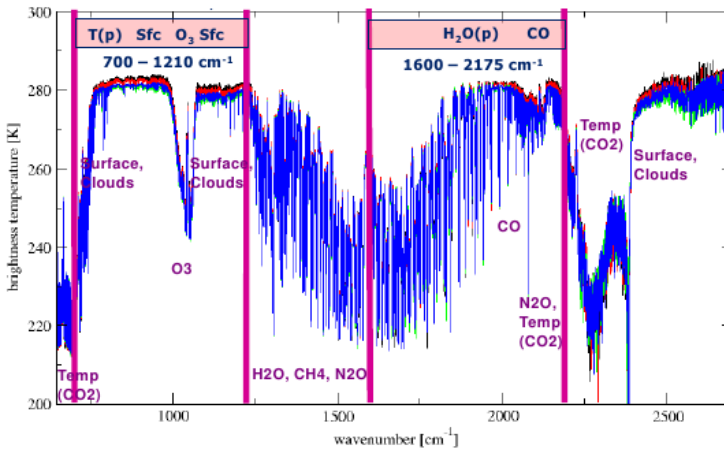
radiometric noise levels could affect the exploitation of the information contained in these spectral channels. Concerning the sounding requirements on spectral resolution, the interferograms taken by the FTS will be sampled based on fixed increments of the translating mirror within the interferometer. When those interferogram samples are Fourier-transformed and calibrated, a spectrum of the incoming radiation is produced with a spectral sampling ( $\Delta\nu$ ) determined by the MOPD of the interferometer's scan mirror, according to the formula:  $\Delta\nu = 1/(2 * MOPD)$  .

<sup>4</sup>The original requirement for the scanning strategy was structured in sequences repeated every 6 hours according to the following patterns: 5 times (LAC3 + LAC4) followed by 4 times (LAC2 + LAC4) followed by 3 times (LAC1 + LAC4).

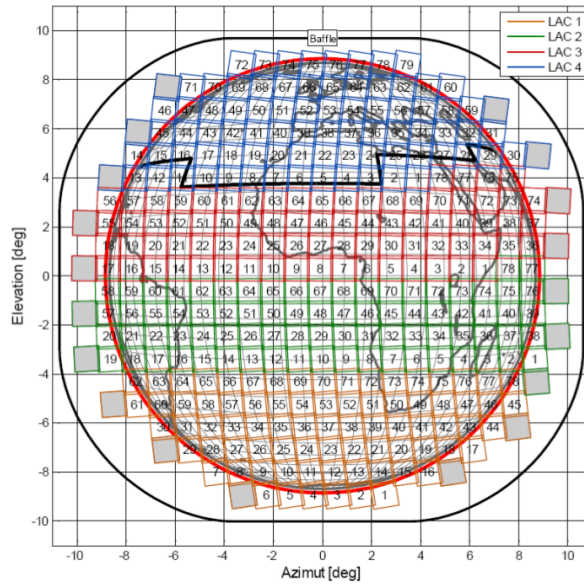


**Figure 2.6:** Illustration of the IRS instrument (preliminary design of Kayser-Threde GmbH).

Processing Facility (IDPF), where they are converted into Level 1b radiance spectra and passed to the Level 2 Processing Facility (L2PF). Both L1 and L2 data are finally disseminated to the end users.



**Figure 2.7:** Spectral coverage of MTG-IRS, identifying the two bands of the instrument, compared to METOP-IASI coverage spanning the whole wavenumber range ( $645 - 2760 \text{ cm}^{-1}$ ) shown in the figure (image credit: EUMETSAT).



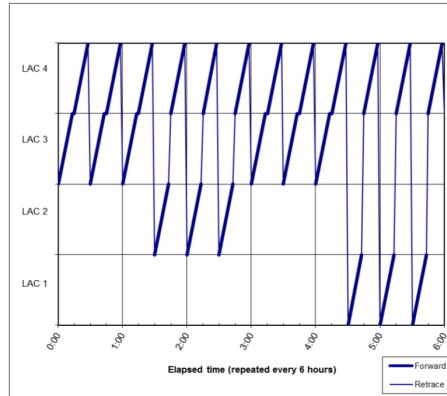
**Figure 2.8:** Local Area Coverage scanning zones, as from [6] (image credit: EUMETSAT).



Note: The diagram indicates the mandatory LAC zone 4 coverage in terms of latitude and longitude on the earth.

AB	30°N
BC	40°W
CD	22°N
DE	15°E
EF	30°N
FG	50°E
GH	35°N

**Figure 2.9:** Details of LAC zone 4 (image credit: EUMETSAT). According to [6], the polygonal line AH defines the mandatory lower boundary of LAC4.



**Figure 2.10:** Operational practice for the IRS scanning sequence, as from [6] (image credit: EUMETSAT).

### 2.3.2 User requirements and expected benefits

As highlighted in the previous section, the IRS (also referred to as GeoSounder) objectives are to provide breakthrough measurements from space of the Earth-atmosphere system. The instrument and scanning specifications are targeted in particular to the detection (signal extraction) and the tracking (time evolution) of horizontal and vertical water vapour structures in the atmosphere. Having no heritage instrument in previous Meteosat missions, this is of course an unprecedented source of information available for the operational nowcasting (NWC) services exploited by the National Meteorological Services (NMS), as well as for the assimilation into atmospheric models performed by the regional and global Numerical Weather Prediction (NWP) Centres.

Whatever the application user, the IRS observations will be extremely beneficial for monitoring the advection and the convergence of low-level moisture in pre-convective environment, often associated with severe weather phenomena as heavy thunderstorms, squall lines and Mesoscale Convective Systems (MCS).

As indicated in section 1.2, data from polar-orbiting IR hyperspectral sounders (IASI on METOP, CrIS on Suomi-NPP) are routinely assimilated into NWP models, to characterise air masses and clouds and to monitor the atmospheric instability. The incomplete exploitation of such hyperspectral measurements in the fields of Nowcasting and Very Short Range Forecast is related to two main limiting factors, both arising from the polar orbit, namely the small number of overpasses per day and the timeliness<sup>5</sup> issues.

<sup>5</sup>According to the definition provided in the EUMETSAT MTG Conventions & Terms



Variable	Accuracy	Horizontal resolution	Vertical resolution	Temporal resolution	Conditions	Priority	Breakthrough
Specific Humidity profile	10 % - 5 %	200 – 3 km	2 – 0.1 km	12 – 0.5 h	TG, CL	VH	dx =30 km, dz = 1 km, dh =3h.
Temperature Profile	1.5 – 0.5 K	200 – 3 km	2 – 0.1 km	12 – 0.5 h	TG, CL	VH	dx = 30 km, dz = 1km, dh = 1h

Notes:
Accuracy is reported as Threshold – Goal values
Horizontal resolution (dx) is reported as Threshold – Goal values
Vertical resolution (dz) is reported as Threshold – Goal values
Temporal resolution (dh) is reported as Threshold – Goal values
Conditions: TG= possible targeted, CL = in, outside and below clouds
Priority level: VH = Very High

**Figure 2.11:** High level user requirements for the level 2 products derived from MTG-IRS observations (from [6]).

To overcome the polar orbit constraints, geostationary hyperspectral sounder would provide, at 30 minutes repetition cycle for the extended Euro - Mediterranean domain (LAC4), accurate Level 1 spectra of absorption lines correlated to low level moisture gradients, low level inversions, tropopause inversions, commonly referred as key factors for triggering severe storms. Moreover, derived Level 2 products, expressed as geophysical quantities, would provide the end users with an equivalent and lower dimensionality information on 3D humidity and temperature structures, surface properties and cloud properties.

To ensure the target performance of the IRS instrument for meteorological applications, the so-called high level user requirements for Level 2 products were established through an elaborate user consultation process, resulting in dedicated position papers prepared by authoritative representatives of the communities Global-NWP, Regional-NWP and NWC. As an example, in the figure 2.11 the high level user requirements for temperature and humidity profiles in cloud-free areas as needed for regional scale NWP applications are reported, extracted from MTG-EURD document [6].

Finally, it is worth to remark that, even though the main mission objective of the MTG-IRS instrument is to provide high spatial and temporal information of the atmospheric thermodynamical state in support of operational NWC and NWP applications, its sounding capability is designed to provide, by side, accurate air composition information (e.g. dust, volcanic ash, trace gases), in support to state-of-art emerging operational atmospheric air quality and air chemistry applications [1].

---

Document (EUM/MTG/DEF/08/0034), the timeliness is the time difference between foreseen end of acquisition of the last contributing data, and the end of reception by the user.



# 3

## CONCEPTS AND METHODS

---

*This chapter describes the scientific environment and the preliminary experiments on the use of IR hyperspectral soundings for nowcasting applications, introducing the theoretical concepts and the experimental methods used throughout the thesis.*

### Contents

---

3.1	The potential of IR hyperspectral sounding . . . . .	22
3.2	IASI interferometer as proxy for IRS . . . . .	26
3.3	The recipe of this project . . . . .	29

---

### 3.1 The potential of IR hyperspectral sounding

As highlighted in the previous chapter, hyperspectral sounding<sup>1</sup> has in principle a great potential for nowcasting applications. This section provides further elements in order to better define the problem to be investigated and to evaluate the impact of such observations in terms of meteorological skill scores.

As reported in [2], nowcasting (NWC) is carried out in regional or national weather forecasting centres by extrapolation of observations, using heuristic rules to modify these observations into the future as well as Numerical Weather Prediction models (NWP) forecasts, with temporal scales from zero to 2 hours ahead. Very Short Range Forecasting (VSRF) is defined as the prediction of the future state of the atmosphere up to 12 hours ahead. Depending on the phenomena, NWC and VSRF cover spatial scales from the micro-alpha (hundreds of metres to 2 km) to the meso-alpha (200 – 2000 km). Temporal scales are from a few minutes to 12 hours or more. At the larger end of the spatial and temporal scales, there is a transition to synoptic scale with phenomena such as extratropical and tropical cyclones.

Nowcasting and VSRF techniques were first practised and developed due to the needs of the aeronautical community. Nowadays, they are applied to many phenomena, as (1) convective storms with associated effects; (2) mesoscale features associated with extra-tropical and tropical storms; (3) fog and low clouds; (4) locally forced precipitation events; (5) sand and dust storms; (6) wintertime weather (ice, glazed frost, blizzards, avalanches); (7) wild fires and (8) contaminated areas. In recent years, there is a clear tendency to develop the nowcasting of “severe” weather, which also requires a special operational routine for the issue of warnings. These warnings are based on specific regional needs but also follow specific national or administrative regulations and thresholds.

Of course, nowcasting and VSRF observational requirements are best satisfied by frequent monitoring of the location, intensity, movement and evolution of the phenomena of interest. Important weather elements are: (1) clouds and precipitation; (2) surface meteorological variables of pressure, wind, temperature, moisture, present weather and precipitation accumulation (or snow layer) as well as land cover and structure; (3) 3D wind field; (4) 3D humidity field and (5) 3D temperature field. For each variable listed above is possible to define the observational requirements to meet the ideal performance of the nowcasting/forecasting systems.

---

<sup>1</sup>Where not explicitly indicated, the IR region of the spectrum must be always referred to for hyperspectral sounding in this thesis.

Regarding satellite observations, infrared soundings allow to reconstruct – by solving an inversion problem known as “retrieval” algorithm [10] – the 3D temperature and humidity fields, which are essential to determine atmospheric stability conditions for predicting convective storms initiation, precipitation type, the possibility of hail and lightning. The continuous improvements in terms of spectral and spatial accuracy of satellite sounders is translated in a better quality of the retrieved fields, which enhance therefore the information needed for convective nowcasting [11].

The intrinsic weaknesses of the nowcasting techniques based exclusively on infrared soundings are mostly related, by one side, to the mandatory clear-sky condition, if a complete vertical profiles is needed; partial retrievals, derived for the region above the clouds, may not be sufficient to infer atmospheric instability, often triggered by the thermodynamic state of lower troposphere. By another side, the convective instability is strongly dependent on water vapour patterns, which in turns are affected to a greater extent by the physical uncertainty propagated during the retrieval process. Finally, the 3D-wind field, not directly reconstructed through the inversion schemes, is universally recognised as a key factor for the initiation of severe convection at such spatial and temporal scales; the so called Atmospheric Motion Vectors (AMV) derived from sequences of infrared observations may, at least in principle, fill this gap.

From an operational point of view, trying to address a procedural approach to the nowcasting problem – hereafter restricted to the atmospheric instability analysis for early warning of severe convection – the findings of the precursor projects undertaken in the framework of COST-78 initiative<sup>2</sup>, specifically dedicated to the “Improvement of nowcasting techniques”, can be assumed as a reference of this thesis work [12].

The topics covered within COST-78 emphasize the climatology of the convection systems over the Euro-Mediterranean domain, expressed through the identification of the relevant patterns of synoptic and mesoscale subjects. Such patterns are often directly related to physical observables measured from remote sensing systems as weather radar, lightning detection networks and, from space, by geostationary (GEO) and polar (LEO) satellites. Another interesting field of research concern the evaluation and adaptation, in some cases, of proper instability indices as predictors of convective systems, spanning multiple spatial and temporal scales. Such predictors are identified as good potential candidates to be processed by automatic detection systems<sup>3</sup>. To some extent, it is easy nowadays to catch a close likeness to the current so-called “expert systems”,

---

<sup>2</sup>COST actions are funded by the European Commission to promote the cooperation in the field of scientific and technical research.

<sup>3</sup>The term automatic is used here to indicate the ability by an appropriate and customised software to detect meteorological subjects, in opposition to the classical diagnostic process performed by the meteorologists and known as “analysis”.

developed in the huge framework of Machine Learning (ML). Furtherly, an interesting comprehensive analysis of similar concepts specifically assessed over the Mediterranean scenario can be found in the synopsis [13].

More recent studies confirm the scientific value of the way ahead drawn by the COST-78 propaedeutic projects. As pointed out in [14], the use of temperature and moisture profiles from AIRS sounder on board NASA Aqua satellite can improve short-term forecasts of sensible weather parameters, due to the ability to resolve some small-scale vertical features coming from the hyperspectral nature of AIRS sounder, thus providing similar quality of the conventional radiosounding profiles. Moreover, in [15] the impact of AIRS temperature and humidity profiles to improve the analysis of a tropical cyclone vortex and the forecast of its rapid intensification is investigated. The findings of the study demonstrate that full 3D spatial resolution of hyperspectral soundings has a positive effect on forecasting the rapid development of tropical cyclones.

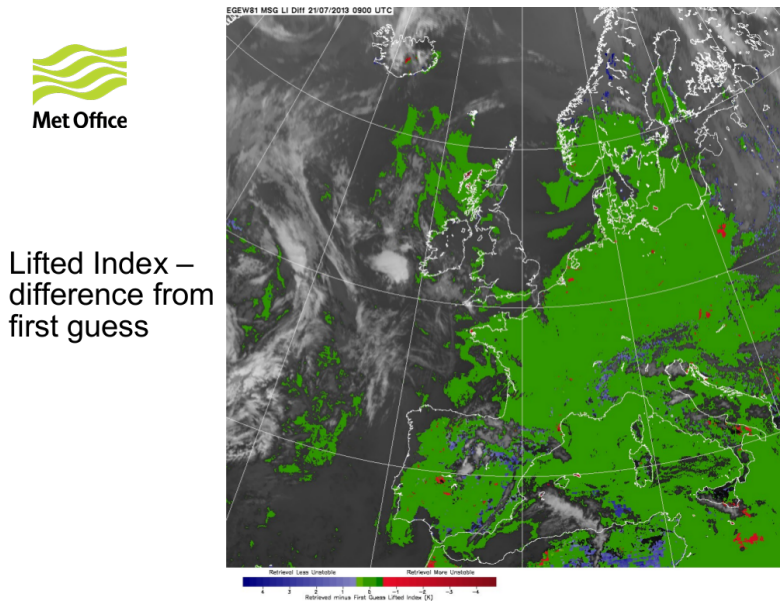
In the context of the EUMETSAT MTG-IRS preparatory studies, conducted in 2009-2013 within the team of experts nominated by EUMETSAT and named MTG-IRS Science Team (MIST)<sup>4</sup>, involving the author as member, highest priority took the investigation of the impact of Level 2 on both NWC/VSRF and NWP activities. Dedicated sessions of the MIST work analysed the potential of future hyperspectral soundings from geostationary, high-rate scans looking at the benefits for application users.

As example of such studies, at the UK Meteorological Office (UKMO) the Lifted Index instability index derived from METOP IASI hyperspectral soundings was used in comparison with the analogous output of the very high resolution NWP model to detect, by difference, the potential areas of poor quality of the NWP forecast (see figure 3.1).

Another area of investigation of MIST addressed the impact of the increased vertical resolution of temperature and humidity profiles from current and future hyperspectral sounders on the detection of atmospheric instability in pre-convective environment (see figure 3.2). As pointed out in [16] and subsequent works, the source of the better vertical resolution arise from the larger information content of hyperspectral data, also referred to as Degrees of Freedom of Signal (DFS), producing a higher number of independent layers. From a theoretical point of view, improvements of atmospheric analysis are expected due to more independent moisture layers localised closer to Earth's surface as well as down to cloud top. Furtherly, similar increased quality of the temperature profiles positively affect the diagnosis of thermal inversions and

---

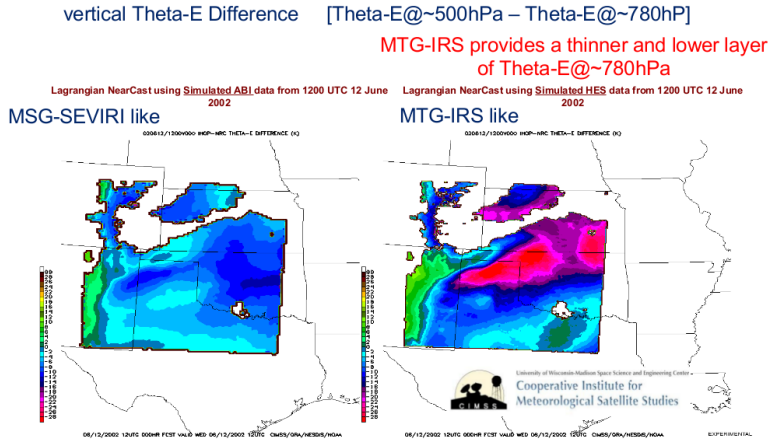
<sup>4</sup>The MTG-IRS Mission Advisory Group (MIMAG) was nominated in 2014 to continue the work initiated by MIST with renewed Terms of Reference (ToR) including advisory responsibility.



**Figure 3.1:** Example of use of satellite-derived instability index for nowcasting purposes at UKMO (image courtesy of Ed Pavelin).

tropopause height. Finally, better soundings allow more accurate computation of Convective Available Potential Energy (CAPE) index<sup>5</sup> and can be used as an independent validation dataset to monitor performance and deficiencies of NWP models.

<sup>5</sup>Theoretical aspects about atmospheric thermodynamics and convection can be found in [17] and [18].



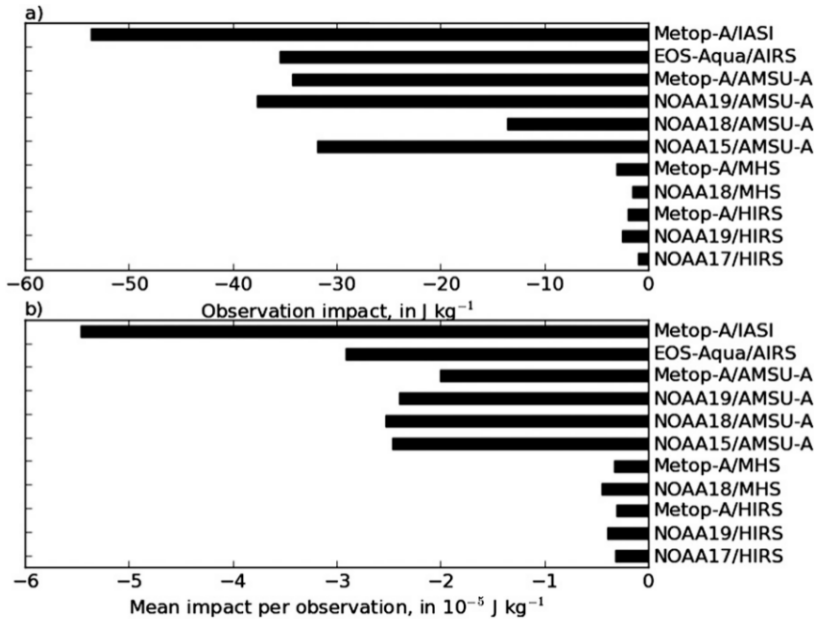
**Figure 3.2:** Example of comparison between MSG-SEVIRI to MTG-IRS for the detection of instability at CIMSS (image courtesy of Ralph Petersen).

## 3.2 IASI interferometer as proxy for IRS

Preliminary considerations reported in the previous section lead to the opportunity to identify the IASI hyperspectral interferometer on board EPS METOP satellites as potential precursor, even though in polar orbit, of future MTG-IRS sounder. As matter of fact, the information content caught by this instrument is awesome, as demonstrated by sensitivity studies carried out at UKMO, focusing on the assessment of the WMO Global Observing System (GOS) [19]. The role of observations in reducing 24-h forecast errors was evaluated using the adjoint-based Forecast Sensitivity to Observations (FSO) method developed within the UKMO global numerical weather prediction system. Satellite data were found to account for 64% of the short-range global forecast error reduction, with the remaining 36% coming from the assimilation of surface-based observation types. Among the satellite observations, microwave and hyperspectral infrared sounding systems were found to give the largest total impacts. The figure 3.3 shows the observation impact and mean impact per observation of each sounder analysed at UKMO. The impacts per sounding of the hyperspectral IR sounders, MetOp-A/IASI and Aqua/AIRS, are the largest, followed by the microwave sounder AMSU-A on board both EPS and NOAA platforms.

With more than ten years of continuous service, very high reliability and



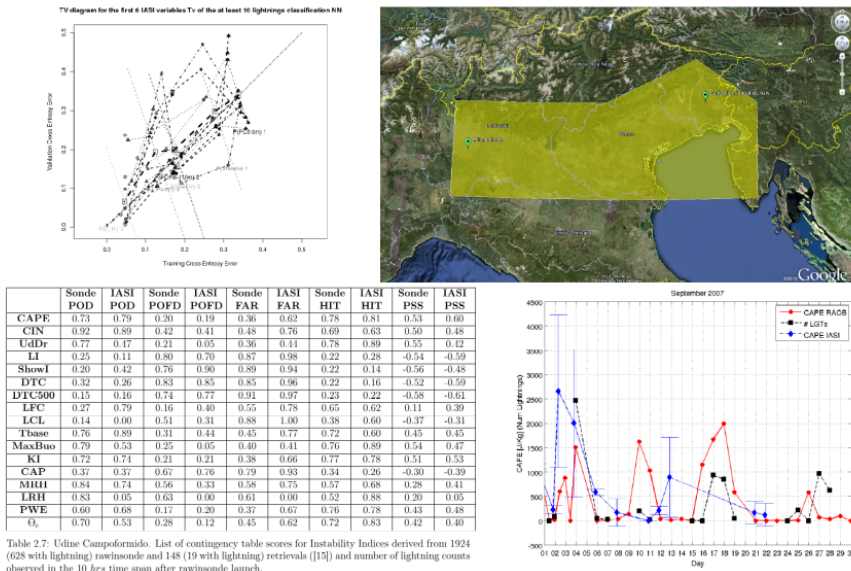


**Figure 3.3:** (a) Observation impact and (b) mean impact per observation of the instruments using MWS and IRS techniques (from [19]).

efficient data delivery and retrieval architecture, IASI represents a pillar of the hyperspectral operational and research communities. In this perspective, we could consider IASI as good “proxy” for IRS.

Based on the findings of MIST studies, a dedicated project funded by EU-METSAT started in 2010 to evaluate atmospheric instability from IASI over Italy, whose the present author was charged as external supervisor. The final technical report of the project [20] highlights the great potential of IASI Level 2 (vertical profiles of temperature and humidity) and Level 3 (instability indices) in order to define accurate and reliable predictors to train and exploit an expert system (in ML sense) for the detection and forecast of thermo-convective thunderstorms. The testbed of the experiment, the table of instability indices evaluated and some plots are reported in figure 3.4.

In late 2012, the research topic of this thesis was conceived and started. Propaedeutic experiments involved IASI Level 2 data available operationally at the Centro Nazionale di Meteorologia e Climatologia Aeronautica (CNMCA, renamed COMET as of today) located in Pratica di Mare (Rome, Italy). As first objective, the research aimed at the reconstruction of IASI thermodynamic profiles and the computation of derived instability indices to be evaluated with respect to the initiation (named also “triggering”) of convection.



**Figure 3.4:** Prediction of instability over Italy using IASI data (image courtesy of Paolo Antonelli and Agostino Manzato). Training-Validation (TV) diagram of the Cross-Entropy Errors (CEEs) computed over multiple instances of training/validation sets to infer the prediction skill of the Neural Networks needed to select the best predictors of the classification algorithm (top-left). Region of interest for the study (top-right). Contingency table scores for Instability Indices derived from radiosoundings and IASI retrievals compared with lightning occurrences (bottom-left). Time series of the CAPE indices computed from radiosoundings and from IASI retrievals, overlapped with lightning coincident events (bottom-right).

---

### 3.3 The recipe of this project

---

As explained in the previous sections, the analysis of the pre-convective tropospheric environment by means of modern hyperspectral sounders observations – such as current IASI on board METOP-A/B and CrIS on Suomi NPP, and future IRS on MTG – plays a crucial role to characterize the atmospheric instability for nowcasting and very-short range forecast activities. In particular, hyperspectral Level 2 products are able to provide added value for retrieving information on horizontal and vertical gradients of moisture and temperature, known as decisive factors for the initiation of severe weather events. Aiming at the early detection of convective systems, this research thesis deals with the setup of an expert system framework, making use of either real (observed, IASI/CrIS) or simulated (proxy) hyperspectral data, as well as of auxiliary colocated data, coming from observations and NWP model data available at the Italian Air Force Meteorological Centre (COMET, formerly CNMCA). The final goal is to assess the correlation between the signal (the “signature”, i.e. the information content of level 2 products) and the phenomenon (the triggering of convective instability).

As scientific baseline of the algorithms exploited throughout the research, background knowledge must be assumed. As first assumption, the use of derived atmospheric stability indices such as convective available potential energy (CAPE) and lifted index from advanced IR soundings can provide strong information on the potential development of severe convective storms within typical NWC and VSRF time scales. This topic is widely investigated in [21]. A second fixed point underpinning this thesis is the possibility of analysing the hyperspectral data by applying a large variety of methods taken from the signal processing theory or, more in general, from the Machine Learning world. As reference, extensive treatment by [22] for the optimal information extraction from multispectral and hyperspectral spaceborne data is hereafter considered. It is worth to highlight, to this regards, that the nature of the problem (the early detection of strong convection) intrinsically brings a geometric fingerprint which can be easily extracted by so called “clustering algorithms”, able to identify pixels aggregated under a proper metrics. An example of such feature extraction has been implemented for the case study shown in section 6.3.

Besides the baseline algorithms, for the identification of significant case studies, as well as for the implementation of any ML method, the signature of severe/hazardous weather needs a strict definition. For the convective instability

characterisation, diagnosing and monitoring and relative definitions, the paper prepared by the EUMETSAT Convection Working Group [11], concerning the best practices on this subject, is the reference<sup>6</sup>.

The baseline data processing include NWP system, observations and auxiliary<sup>7</sup> data. Regarding the NWP component the coupled atmospheric analysis-forecast system, based on CNMCA data assimilation system and COSMO forecast model, is used (see figure 3.5). The observations, to be correlated to the triggering of convective systems, may be conceptually divided in three classes:

- hyperspectral IR products (e.g. level2 temperature and humidity profiles), independent from NWP state and carrying the signal to be detected;
- non conventional “4D/high density” datasets (e.g. lightnings, rain-gauges, weather radars, etc.), independent from NWP state and acting as reference for calibration/validation of the processor;
- conventional observations (e.g. radiosoundings, airborne data), regularly assimilated in NWP state and acting as control data.

Concerning hyperspectral data, in parallel with real spaceborne data the Level 2 products generated in near real time mode by EUMETSAT for MIMAG activities, as proxy for IRS, are considered highly beneficial (see next chapter). Spatial and temporal resolutions for all observing datasets are accorded to the nowcasting and very short range forecast scales:

- time: sub-hourly to hourly;
- space: micro-alpha (hundreds of meters) to meso-alpha (hundreds of km) scale.

Finally, considering the size and the timeliness for future IRS level 1 data and recognising that the use of level1 – for Meteorological Centres with medium/low ICT capacity – is a process strongly resource and time consuming (especially in on-line mode), the best exploitation of the hyperspectral IR L2 products is a strategic keypoint, to demonstrate the end-user readiness on day1.

---

<sup>6</sup>Just as an example, the term “pre-convective environment” is referred to the 4-D thermodynamic and wind field present before convective initiation (CI) occurs.

<sup>7</sup>Auxiliary data include e.g. independent calibration/validation datasets, used to compensate the possible lack of a priori 3D information content in current data assimilation systems, resulting in deficiencies of NWP forecasts. As examples of such category, products from EUMETSAT CAF (GII product), NWC SAF (cloud classifiers, RGB air mass, physical retrievals) and Hydrology SAF (precipitation products as convective rain intensity) are taken into account, helping to identify the significant “patterns” of severe weather.

**CNMC NWP system overview**

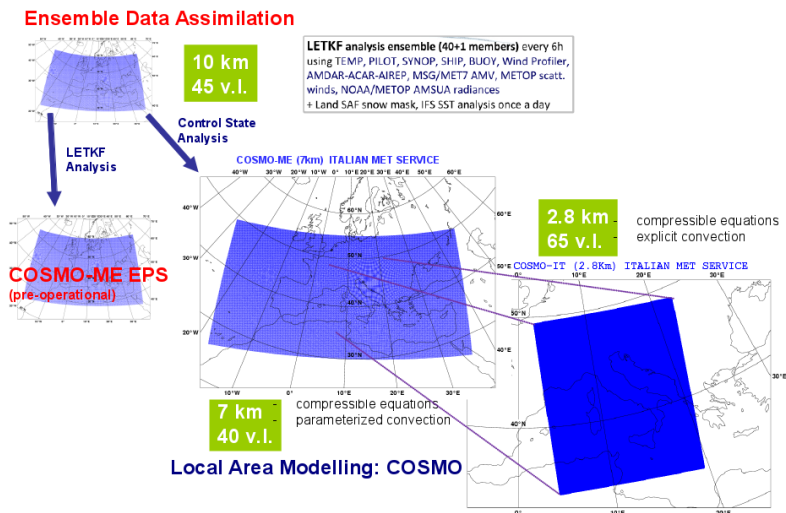
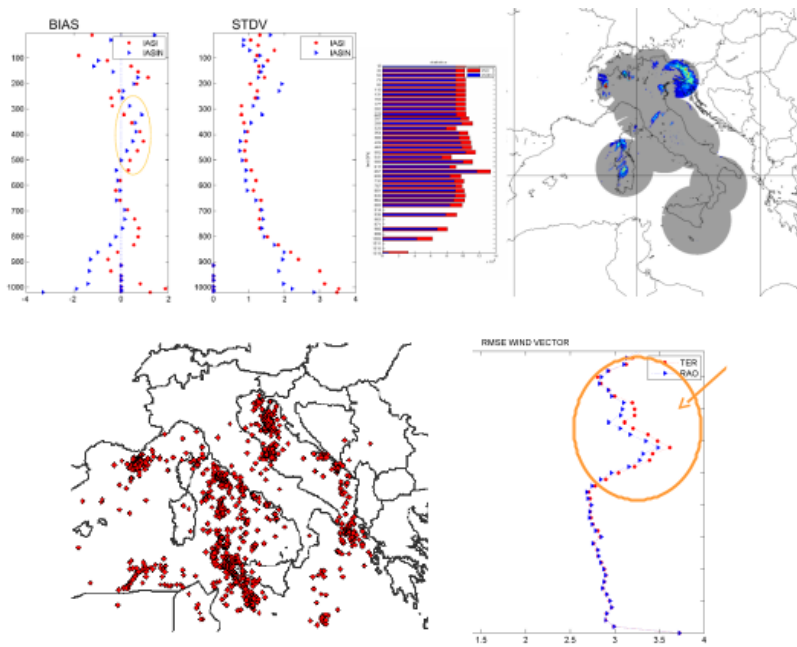


Figure 3.5: Baseline Data Processing – NWP component.



**Figure 3.6:** Baseline Data Processing – observation component. IASI (top-left) and radiosounding (bottom-right) assimilated profiles, with associated impact on objective analysis. Italian weather radar (top-right) and lightning (bottom-left) networks.

# 4

## THE MTG-IRS PROXY DATA ITALIAN DEMONSTRATION PROJECT

---

*This chapter provides a description of the demonstration project conceived and realised by the author in support of the EUMETSAT MTG-IRS Mission Advisory Group activities.*

### Contents

---

4.1	EUMETSAT IRS NRT demonstration project . . . . .	34
4.2	ItDP software package . . . . .	36
4.3	An example of IASI data processing . . . . .	38
4.4	Examples of IRS proxy data processing . . . . .	42

---

## 4.1 EUMETSAT IRS NRT demonstration project

---

In the framework of the Phase C/D research activities aiming at the day-1 user preparedness, to further explore the potential of MTG-IRS products for nowcasting operational applications, EUMETSAT proposed in late 2015 to consider level 2 products derived from current real observations, instead of relying on historical cases as done before. This action was in line with a specific recommendation arising from the first MTG-IRS Nowcasting Workshop (Darmstadt, 25-26 July 2013). To successfully investigate the potential of the level 2 products for day-1 applications a technology demonstrator of Near Real Time (NRT) services of level 2 products derived from MTG-IRS proxy data, starting from May/June 2016 was set up. The generation of level 2 products in NRT mode was based on actual observations (raw radiances) of IASI on board Metop-A/B and, potentially, of CrIS on Suomi-NPP, considered both as IRS proxy data. The participation of a set of selected scientific investigators through an application proposal was, of course, a key point of the project, involving the operational community at an early stage of development of the MTG-IRS processing facilities. During the NRT-Demo project, started in mid-2016 and lasting one year, level 2 products were generated at EUMETSAT premises using the Level 2 Validation and Demonstration Processor (L2VDP) and distributed in near real time to the participants using the EUMETSAT terrestrial dissemination system (EUMETCAST) [23].

L2VDP is an end-to-end processor of observations by hyperspectral infrared sounders. Given an observation, it classifies the observation according to the presence of clouds, generates level 2 products like temperature and humidity profiles and performs a transformation of these regular products into products specific for data assimilation applications. The processing sequence is completed by a quality control step and the generation of different output files in the NetCDF-4 format. The processor has been developed in support of EUMETSAT's MTG-IRS programme through a concerted effort and dedication of a large number of scientists and software engineers. As theoretical baseline, L2VDP performs a physical (Bayesian) retrieval starting from raw hyperspectral data (Level 0) of actual or historical hyperspectral observations made by IASI or CrIS, considered as proxy. Regarding the volume of data processed, the numbers involved are the following:

**IASI:** a granule represents 3 minutes of observations made by IASI, resulting in a total of 2760 measurements (FOVs) per granule. There are 480 granule per day, resulting in 1.3248.00 observations per day.



**CrIS:** a granule represents 32 sec of observations made by the CrIS instrument resulting in 1080 measurements observations per granule. There are 2700 granule per day, resulting in a 2.916.000 observations per day.

Assuming a global cloud cover of 50% per day about 2.782.000 observations available are expected for processing. Last, L2VDP generates approximately 3200 products per observation (as profiles of temperature, humidity and the “scaled projected states”, a special product for data assimilation applications).

MTG-IRS NRT-Demo project was designed keeping in mind the need to work with the users to (1) prepare the target user community for the new class of data, being IRS a radical new mission, and (2) optimise the envisaged products during the development of mission, so that data and products could immediately be used at day-1 operations. Taking advantage of such opportunity, in November 2015 the author submitted an application proposal of the project named “MTG-IRS proxy data Italian Demonstration Project” (ItDP), with the dual objective of (1) demonstrating the potential of the information content of hyperspectral sounding instruments to improve the air mass characterisation in the pre-convective environment, and of (2) enhancing the capacity and user-readiness of modern, operational Meteorological Services (MSs) with respect to early detection of severe weather. The project, based on the experiments and concepts developed until then, as reported in the previous chapter, would consolidate the processing tools, through the setup of a pseudo-operational environment at COMET, providing at the same a sort of Proof of Concept (PoC) for this research topic.

Using a phased approach, the first target of the project was the assessment of the correlation between the signal (hyperspectral signature, i.e. the information content of level 2 products) and the phenomenon (triggering of convective instability, defined through a set of parameters’ thresholds and spatial features taken from literature). Then, based on the findings of the first phase, the project focused on the preparation of an online processor, whose target was the automatic production in NRT of a set of scores and physical variables to support NWC and VSRF applications. After the first year of IRS NRT Demo and ItDP projects, an important review meeting took place in May 2017 at EUMETSAT headquarters, to gather the main findings of the applicant researchers and to address necessary improvements of the L2VDP processor for future investigations.

A relevant component of the project effort involved the creation of a dedicated software package. To accomplish this, a set of software modules for both online and offline applications has been developed from scratch, taking into account the state of art features such as modularity, portability, interoperability, scalability.

All operational processing chains were running at COMET Data Centre, exploiting the maximum sustainable benefits from its satellite, telecommunication, network and High Performance Computing (HPC) facilities.

## 4.2 ItDP software package

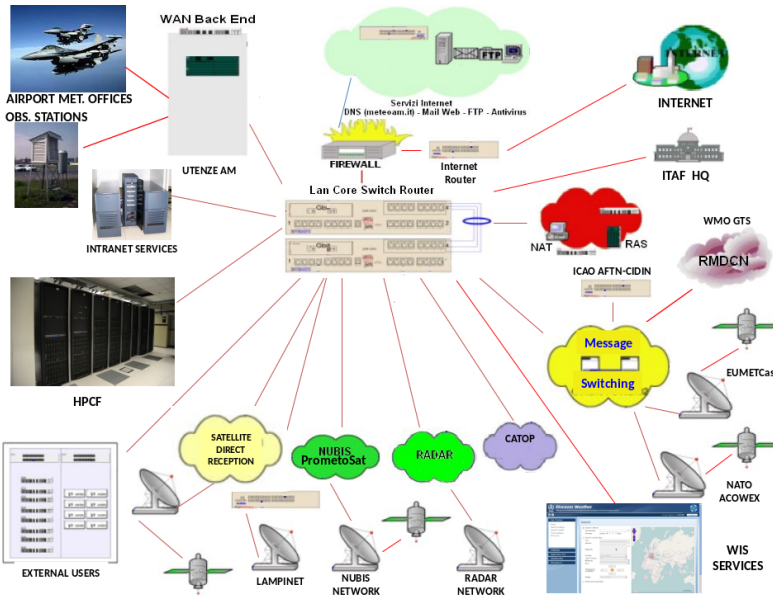
---

As original contribution of the author to this thesis work and to EUMETSAT IRS NRT Demo Project a wide collection of software modules has been designed, tested and implemented on the COMET facilities. All the code of this toolkit uses the state of art, free and open-source programming environments and 3rd party libraries, to ensure to the maximum extent portability, maintenance, sharing and further evolution. In this perspective, the choice of *python* for programming language as well as for standard libraries is considered highly beneficial. All the software runs on *linux* platforms and is maintained and developed under version-control using the efficient tool provided by *git*<sup>1</sup>.

The timely and reliable data flow is ensured by the ICT backbone of the COMET Data Centre, including the redundant network and telecommunication connection with EUMETSAT to receive either in active mode (pull) or in passive mode (push) NRT data.

---

<sup>1</sup>GIT is a Version-Control System (VCS) for tracking changes in computer files and coordinating work on those files among multiple people.



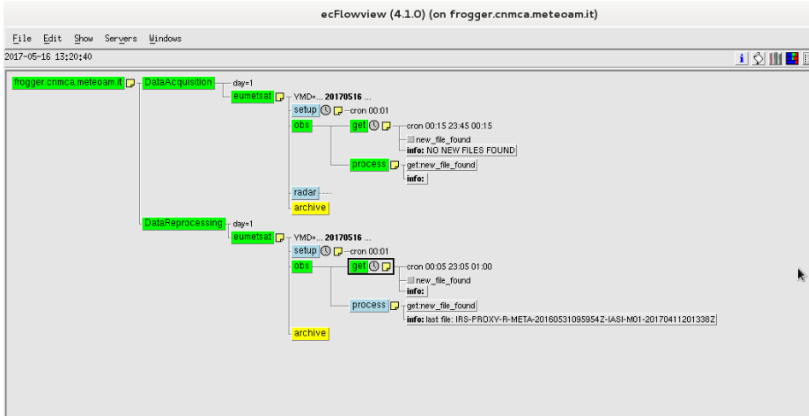
**Figure 4.1:** Schematic picture of ICT assets for data processing and distribution of the COMET Centre.

In detail, the actual IASI/CrIS observations are received operationally with the highest reliability and best timeliness using EUMETCAST<sup>2</sup> service circuits. On the other side, the IRS NRT proxy data generated by L2VDP are downloaded from a dedicated ftp server made available by EUMETSAT specifically for this project.

ItDP software tools can be run either in stand-alone mode or embedded in job schedulers, as needed for example in cascading processing chains also referred as “suites” operated by the meteorological centres. The latter approach has been preferred, especially in the final phase of the research project when large datasets of proxy data were available. To this regard, the scheduler of ItDP suite has been configured in twin mode, namely for the “online” processor, used for operations in NRT, and the “offline” processor, activated on demand for research case studies and for reprocessing experiments in case of new or updated datasets available.

The job scheduler environment chosen for ItDP is based on the free and open-source software *ecFlow* developed and maintained by the European Centre for Medium range Weather Forecasts (ECMWF), widely used at COMET as well

<sup>2</sup>EUMETCAST is EUMETSAT’s primary dissemination mechanism for the near real-time delivery of satellite data and products, based on multicast technology. It uses commercial telecommunication geostationary satellites using Digital Video Broadcasting (DVB) standards and research networks to multi-cast files to a wide user community. It also delivers a range of third-party products, as for example the CrIS data from NASA.



2 parallel processors of data streams:  
NRT (from OPEfix) and REP (from REPfix)

**Figure 4.2:** Job scheduler of ItDP processor for MTG-IRS proxy data, showing both online and offline batch queues.

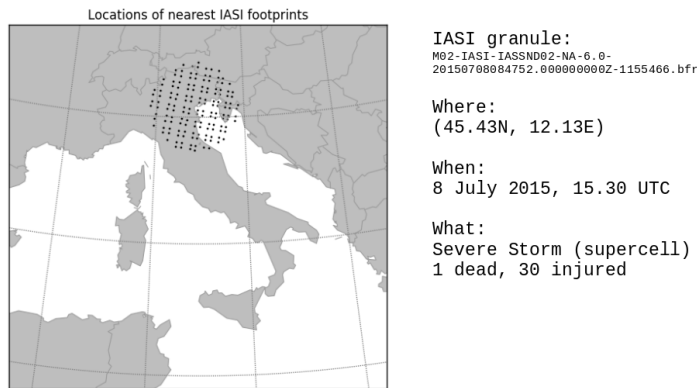
as in many other meteorological services. A snapshot of the ItDP processors controlled by *ecFlow* scheduler is shown in figure 4.2.

To complete the ItDP toolkit, in addition to the software dedicated to IRS proxy data processing other modules for NWP models, weather radar, conventional observations and auxiliary datasets are available. Finally, routines for colocation, reprojection, interpolation, smoothing, mapping, plotting, statistics and Machine Learning are ready to use. More details on the developed procedures can be found in chapter 5.

## 4.3 An example of IASI data processing

As first example of capability of ItDP processor, METOP-A IASI actual data for a severe weather event, striking north-eastern Italy in summer 2015, were retrieved from EUMETSAT archive and analysed. Figure 4.3 shows on the map the locations of the nearest IASI footprints (FOVs) to the event, consisting in a rotating cluster of thunderstorm cells<sup>3</sup>. IASI Level 2 data granule was

<sup>3</sup>This case was associated to an Enhanced Fujita scale-level 4 tornado (see Zanini et al., 2017; <https://link.springer.com/article/10.1007/s11069-017-2741-6>).

*Case Study: 8 July 2015, Venice area*

**Figure 4.3:** Example of IASI historical data downloaded from EUMETSAT EO portal and processed by ItDP processor.

decoded and geolocated, obtaining 3D atmospheric fields of temperature and humidity as well as ancillary metadata (as for example quality indicators and surface parameters), useful to discard bad quality retrievals or apply filters by land/sea qualifiers.

Thermodynamic vertical profiles, named soundings, can be analysed computing for each single column the potential degree of energy available to trigger the convection, expressed in terms of derived instability indices. In figure 4.4 for each FOV is plotted the Convective Available Potential Energy index, demonstrated to be one of the most correlated to the development of strong convection (see for example the comparison among different indices made in [20]).

A closer look to the thermodynamic profiles by means of the classical Skew-T (or Herlofson) diagram reveals full details about the origin of the latent instability measured by CAPE index, as shown in figures 4.5 and 4.6 for two FOVs with highest values of CAPE, whose the second is very close to the event. It is worth to notice, from the nowcasting point of view, that the overpass of METOP-A was more than six hours before!

A simple but efficient tool to analyse the distribution of instability “strength” in the area of interest is of course the histogram plot, shown for this case study in figure 4.7.

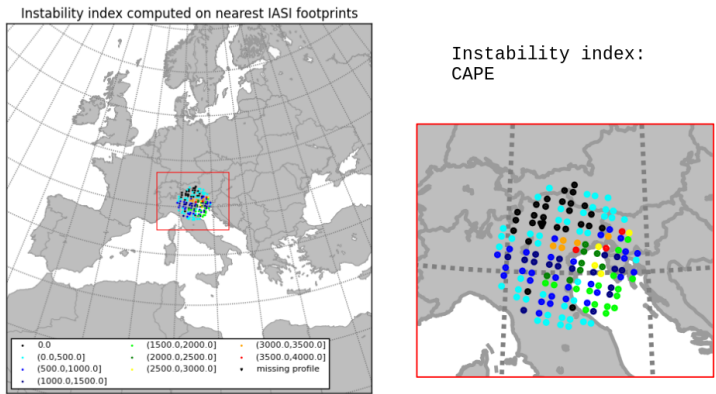


Figure 4.4: Example of map of IASI instability index (CAPE) computed by ItDP processor for a case study.

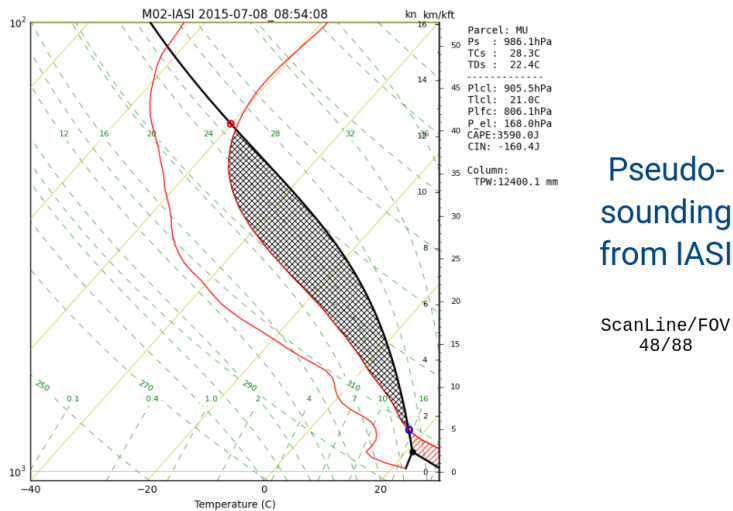
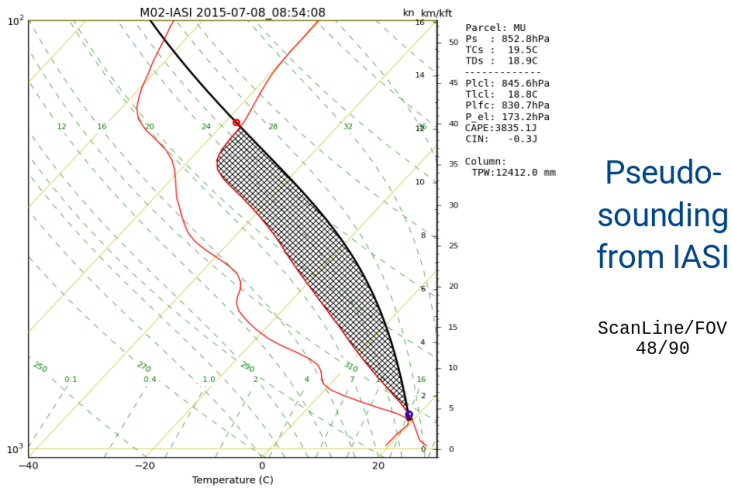
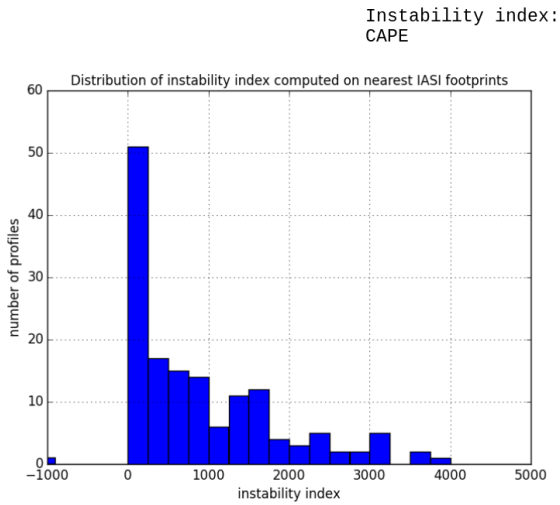


Figure 4.5: Example1 of Skew-T diagram of IASI data computed by ItDP processor.



**Figure 4.6:** Example2 of Skew-T diagram of IASI data computed by ItDP processor.

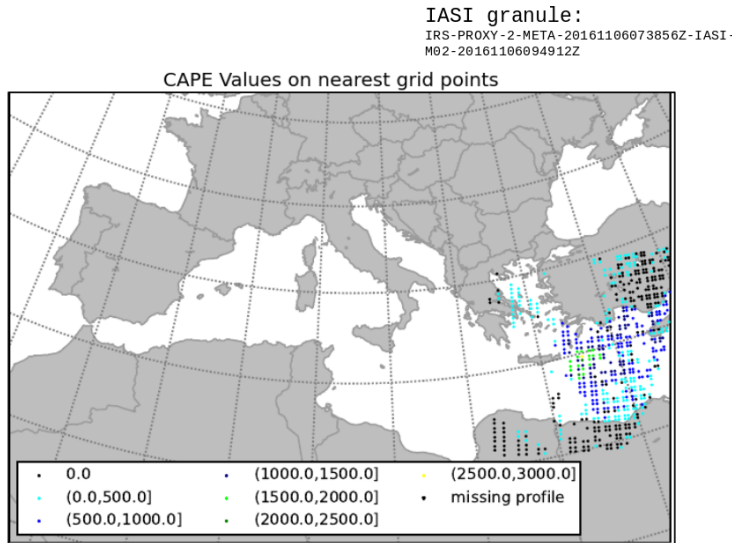


**Figure 4.7:** Example of distribution of IASI instability index (CAPE) computed by ItDP processor.

## 4.4 Examples of IRS proxy data processing

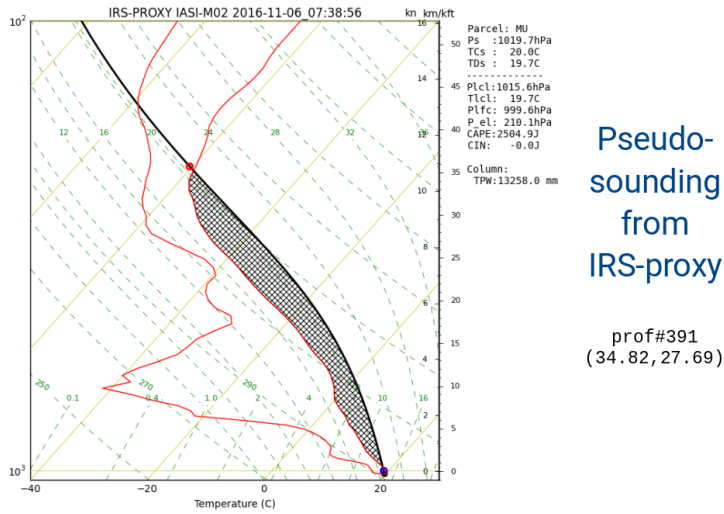
In analogy with the example of IASI of the previous section, IRS proxy data can be fully exploited by ItDP processor to prepare subsequent signal extraction and correlation studies. Both online and offline processors are pretty tunable defining for example the area of interest (as bounding box or radial distance), the instability index and its eventual threshold to restrict the selected cases. Figures 4.8 and 4.9 show, respectively, the mapped CAPE values and a single FOV Skew-T diagram for a IRS proxy data granule derived from IASI and processed in online mode. Just as final note, all the plots and post-processing data are automatically produced by the ItDP software, to confirm the envisaged potential utilisation by the nowcasting user community.

Similarly, figure 4.10 shows the map of CAPE values for multiple, successive IRS proxy data granules derived from CrIS and reprocessed in offline mode by ItDP processor. The higher spatial resolution of CrIS, due to the different scanning features, is evident here.

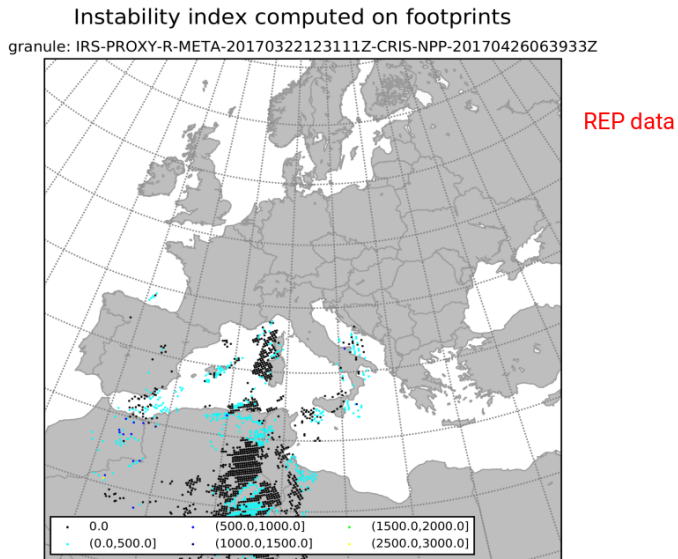


**Figure 4.8:** Example of map of IRS-proxy instability index (CAPE) computed by ItDP online processor.





**Figure 4.9:** Example of Skew-T diagram of IRS-proxy data computed by ItDP online processor.



**Figure 4.10:** Another example (CrIS as input) of map of IRS-proxy instability index (CAPE) computed by ItDP offline processor.



# 5

## PROJECT METHODOLOGY

---

*This chapter provides the exact details on the original procedures designed, implemented and tested to achieve the aim of the project, as baseline methodology for ItDP and upcoming developments.*

### Contents

---

<b>5.1</b>	<b>Software Architecture . . . . .</b>	<b>46</b>
<b>5.2</b>	<b>Functionalities . . . . .</b>	<b>51</b>
5.2.1	Processing System . . . . .	51
5.2.2	Supervisor System . . . . .	56
5.2.3	External Libraries . . . . .	57
<b>5.3</b>	<b>An example of end-to-end processor . . . . .</b>	<b>58</b>
<b>5.4</b>	<b>What's next? . . . . .</b>	<b>60</b>

---

As introduced in the previous chapter, the ItDP toolkit is based on a modular architecture build on top of a collection of processing functions, developed from scratch and implemented on the COMET ICT facilities. According to the standard framework, originating from the manufacturing on an assembly line, the approach referred to as Systems Development Life Cycle<sup>1</sup> (SDLC) has been applied here. The main phases of the cycle include planning, analysis, design, development, testing, implementation and evaluation, sequentially correlated in a circular diagram similar to the classical Plan-Do-Check-Act (PDCA) or Deming cycle introduced in the Quality Management context. All the procedures of the ItDP processor have been planned, realised and maintained within the SDLC baseline.

## 5.1 Software Architecture

As stated in section 4.2, all the code of ItDP toolkit is written in *python* language and runs on *linux* platforms. Translating the principles of SDLC into the software development process, the Version-Control System implemented through *git* has been largely used for tracking changes and code releases. A snapshot of usage of such tool during the testing phase of the processor is shown in figure 5.1.

The modular blocks (or procedures) of the ItDP software package allow to build a cascading end-to-end processing chain, as for example the online processor and its analogue co-worker for reprocessed data (see figure 4.2). At the same time, single procedures can be run in interactive mode for deeper, customised data analysis as typical example of stand-alone usage.

Looking at the top-level procedures of ItDP, the main blocks and their respective functions can be identified as follows:

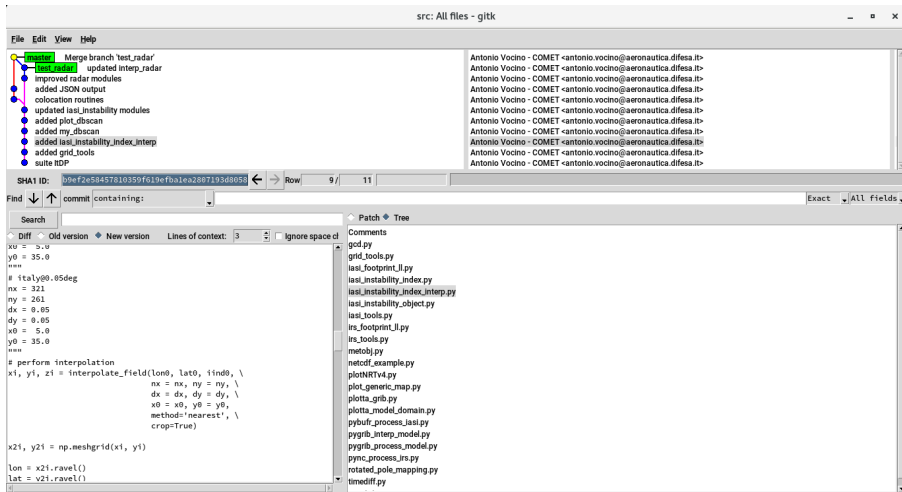
**Grid** data georeferencing, colocation, interpolation, masking, reprojection (as for rotated grid of NWP COSMO model);

**Data Input** decoding of input data in multiple formats (like NetCDF, HDF5, BUFR, GRIB, ASCII), assignment of values to internal data structures;

**Data Output** encoding of output values in many standard formats, encoding of output objects in high-structured format (as JSON);

---

<sup>1</sup>The Systems Development Life Cycle (SDLC), also named application development life cycle, is a term used in systems engineering, information systems and software engineering to describe a process for planning, creating, testing, and deploying an information system.



**Figure 5.1:** Snapshot of the git interface used in this project for track changes and software optimisation.

**Thermodynamics** thermodynamic functions, single-column model, instability indices computation;

**Classifier** data filtering, multivariate analysis, class assignment by structured arrays;

**Object Detection** clustering, assignment of key attributes of meteorological objects;

**Statistics** statistical analysis (as correlation), fitting;

**Quality Control** masking of input data, assignment of quality indicators for output data;

**Map** mapping, output in standard graphical formats (as PNG, PDF) and GIS-oriented formats (as GML, GeoJSON, KML);

**Plot** plotting of cartesian diagrams (as scatter plot, distribution histograms), plotting of thermodynamic diagrams (Skew-T).

The procedures so far may be considered as strictly related to the data processing. In addition, there are the procedures operating at “machine level” to supervise both the computing and the maintenance tasks, namely:

**Data Switching** data download, upload, synchronisation;

**Job Scheduler** scheduling, logging, reprocessing;

**System Maintenance** cleaning, archiving, backup and recovery;

**GUI** supervision, debugging.

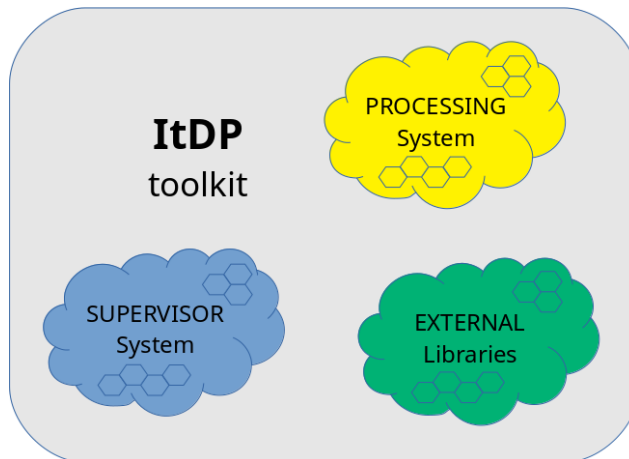
For the sake of completeness, it is worthwhile to include in the list the third party modules. They are native *python* and *C* system libraries:

**Scientific Libraries** scipy, numpy, scikit-learn;

**Mapping and Plotting Libraries** matplotlib, cartopy, proj;

**Decoding and Encoding Libraries** NetCDF, HDF5, BUFR, GRIB (edition 1 and 2), JSON.

The procedures as above, grouped in three classes (clouds) depending on the scope, are sketched in figures from 5.2 to 5.5. Each procedure (indicated in the figure by a hexagon symbol) consists of a set of software modules, expressed as scripts, subroutines and functions, which ultimately represent the elementary building blocks of the whole system. In figure 5.6 an example of such elementary blocks (triangles) for the GRID procedure is shown.



**Figure 5.2:** Functional blocks of the main procedures involved in the ItDP software system.

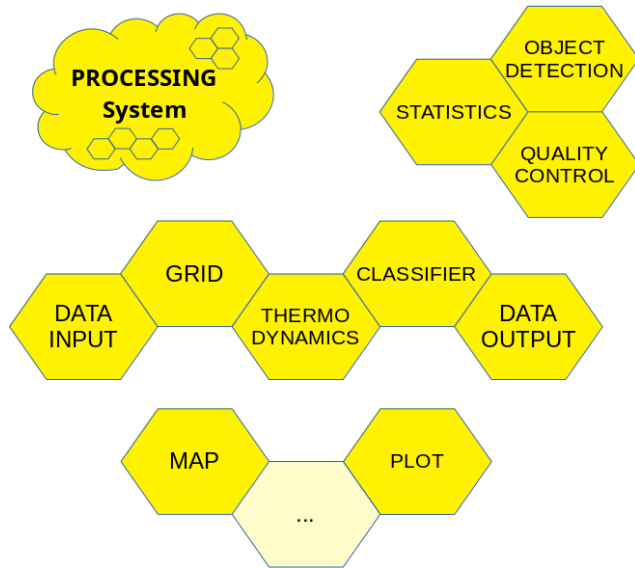


Figure 5.3: Procedures – Processing System.

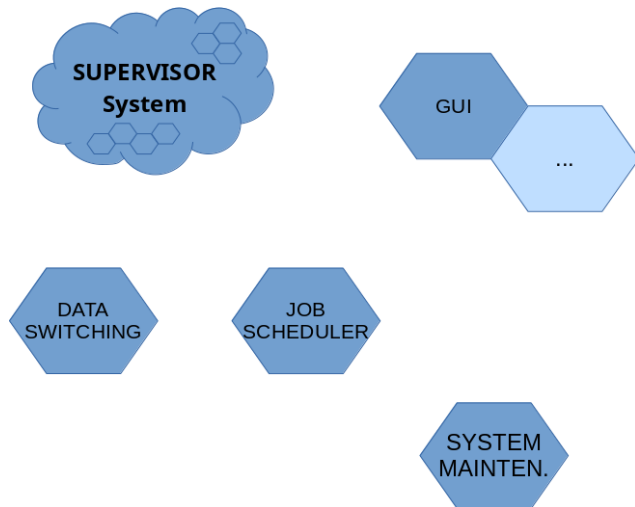


Figure 5.4: Procedures – Supervisor System.

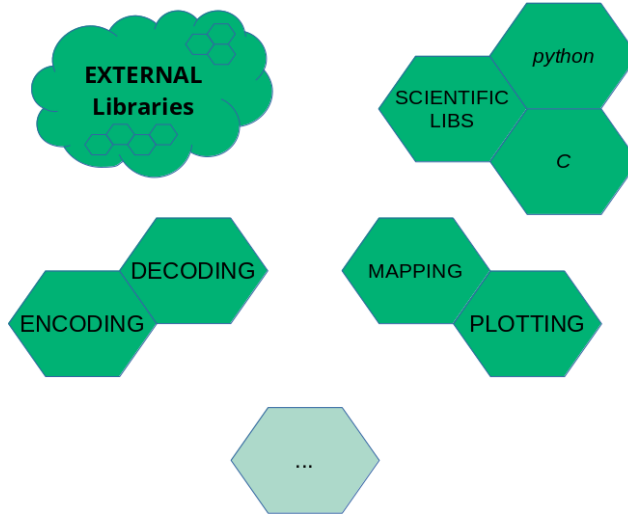


Figure 5.5: Procedures – External Libraries.

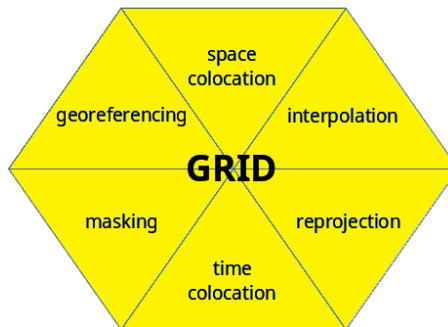


Figure 5.6: Modules of GRID procedure.



---

## 5.2 Functionalities

---

The objective of this section is to provide a deeper description of the functionalities of the various procedures comprised in the ItDP toolkit, in particular for the Processing and Supervisor Systems. The external libraries, as being only a system prerequisite (without any adaptation or customisation needed), are just listed at the end of the section; for any further detail, a link to the software source repository is provided.

---

### 5.2.1 Processing System

---

#### **Grid**

The Grid procedure is a major component of the ItDP suite, designed to perform a series of spatial and temporal operations on the variables involved in all other processors. A typical example of such transformations occur in input phase, when data coming from different sources must be geographically and temporally collocated to a reference grid and timeframe. A very particular case of spatial operator regards the vertical dimension, expressed sometimes in terms of local vertical coordinate, sometimes as pressure coordinate, sometimes in terms of Radiative Transfer Model standard level coordinate. Thermodynamical relationships provided by the thermodynamics procedure ensure the right correspondence. Another customary convention to deal with is related to the rotated grids used by the NWP Limited Area Models to reduce the distortions at mid-latitudes arising from standard (geographical) latitude-longitude coordinates; the reprojection of the atmospheric fields on the reference grid – usually a regular mesh over the area of interest – is therefore needed. Standard projection tools through 3rd party libraries as *proj* have been implemented as auxiliary procedures. Other essential operations provided by the Grid procedure regard the interpolation of the primary variables occurring for example in case of sparse data to be referred to a common (also named “target” hereafter) grid. According to the scientific literature, different algorithms take into account the physical constraints to be applied during the interpolation. As an example, mass-flux conservative requisites suggest to use nearest neighbour method for quantitative precipitation fields, as the Surface Rainfall Intensity (SRI) inferred from the weather radar reflectivity. For smoother fields as NWP mid-tropospheric atmospheric variables, either a simple bilinear or a bicubic interpolation are available. Finally, masking modules allow, at any level, the

selective use of data in dynamic mode, based on quality indicators (as metadata from the Quality Control procedure) as well as on geographical filtering defined by the colocation algorithms. The advantage offered by the masked-arrays structures provided natively by *python* is straightforward.

### Data Input

The Data Input procedure is the interface to the “external world”. To handle the large variety of meteorological data formats, a set of preprocessors has been developed, specific for the different kind of input files. Each module can be considered a front-end of the ItDP toolkit, nested on the system libraries preinstalled for the respective data format (for more details see subsection 5.2.3). The list in table 5.1 summarises the available front-ends of the Data Input procedure, with some further details about the implementation.

**Table 5.1:** List of formats handled by the Data Input procedure.

Format	Note
NetCDF	IRS proxy data, according to the L2VDP data model; input files processed are Atmospheric Temperature Product (ATP), Atmospheric Humidity Product (AHP) and ancillary metadata (META)
BUFR	- IASI data, delivered in NRT via EUMETCAST service; - radar data (single site and Italian mosaic products)
HDF5	radar data (European composite products)
GRIB	NWP model analysis and forecast data, in both edition 1 and 2 versions; ECMWF as global model and COSMO as limited area model are decoded
ASCII	raw input for any other non-conventional data (as for example lightning data from the Italian Air-Force lightning network (LampiNet); Comma Separated Values (CSV) and Tab Separated Values (TSV) files are read

The functional scheme of the Data Input procedure is based on the definition of proper Data Models as abstraction layers, implemented at software level through the utilisation of classes or data structures. The striking feature in such case is the possibility of defining along with the values of the variables also a set of metadata, expressed as attributes or pairs key/value. Metadata can be set at any stage of the use of the classes, for example in input (quality indicators), processing (flags for feature extraction) and output (validation scores). In the following paragraph about Object Detection procedure, a simple prototype of class based on the Data Model designed for instability objects is shown (see figure 5.9). As matter of fact, the baseline of the L2VDP set up by EUMETSAT for IRS proxy is also founded on the use of Data Models, that ultimately are strongly related to the current definition of so-called Atmospheric State Product (ASP) that will be generated by the L2PF from the MTG-IRS

Product Element	Symbol	Unit	Range	Precision	size <sup>1,2,3</sup>	Target	Comment
<b>Navigational Information</b>							
Geographical position	$(\lambda, \phi)$	TBD			2	NRT	Latitude and longitude of each MTG-IRS FOV
Time of observation	h	TBD			1	Near Real Time	
Satellite Zenith and Azimuth	$(\theta_s, \phi_s)$	[1]			2	Near Real Time	
Solar Zenith and Azimuth	$(\theta_\odot, \phi_\odot)$	TBD			2	Near Real Time	
<b>Posterior Atmospheric State<sup>4</sup></b>							
Temperature Profile	$t^{asp}$	K	100 - 350	1	$n_t^{asp}$	Near Real Time	
Specific Humidity Profile	$q^{asp}$	kg/kg	0 - 10	0.001	$n_q^{asp}$	Near Real Time	
Ozone Profile	$o_3^{asp}$	kg/kg	TBD	TBD	$n_{o_3}^{asp}$	Offline	
<b>Posterior Atmospheric State standard deviation</b>							
Temperature	$s_t^{asp}$	K	TBD	TBD	$n_t^{asp}$	Near Real Time	
Specific Humidity	$s_q^{asp}$	kg/kg	TBD	TBD	$n_q^{asp}$	Near Real Time	
Ozone	$s_{o_3}^{asp}$	kg/kg	TBD	TBD	$n_{o_3}^{asp}$	Offline	

**Figure 5.7:** Extract from IRS Level 2 Atmospheric State Product element list (from [9]).

observations at day-1 of operations<sup>2</sup>. For the sake of illustration, an extract of IRS Level 2 Data Model taken from the ATBD reference documentation [9] is reported in figures 5.7 and 5.8.

### Thermodynamics

This procedure provides the comprehensive set of thermodynamic functions necessary to derive the complete information about the atmospheric state on a particular location. Transformations between equivalent thermo-hygrometric variables are performed, based on standard scientific libraries. To evaluate the degree of instability across the troposphere, classical single-column models (as the lifting parcel method, see [18]) have been implemented and interfaced within this procedure. They allow the computation of all the instability indices needed as predictors to detect and track the signal about the convective instability. The figure 4.9 at page 43 is an example of the information made available by the ItDP Thermodynamics procedure.

### Classifier

The objective of the Classifier procedure is to aggregate the data coming from different sources (ingested by Data Input), collocated in space and time

<sup>2</sup>For the IRS mission product extraction chain, day-1 is defined as the date on which the part of the MTG system connected to the first sounding satellite (MTG-S1) is declared operational following its one year commissioning time after the launch of MTG- S1.

Processing Flags							
IRS L1b Quality Test	[-]	[-]	TBD	TBD	1	Near Real Time	IRS observation quality and completeness
Data Acceptance (DA) module check	[-]	[-]	TBD	TBD	1	Near Real Time	DA module applied and exit status.
Product Generation (PG) module check	[-]	[-]	TBD	TBD	1	Near Real Time	1DVAR module applied and exit status.
Post Process (POP) module check	[-]	[-]	TBD	TBD	1	Near Real Time	POP module applied and exit status.
Quality Assurance (QA) module check	[-]	[-]	TBD	TBD	1	Near Real Time	QA module applied and exit status.
Scene Analysis (SCA) module check	[-]	[-]	TBD	TBD	1	Near Real Time	SCA module applied and exit status.
Cloud Test	[-]	[-]	TBD	TBD	1	Near Real Time	Which cloud tests applied and exit status.
Qualitys Flags							
Quality Indicator	[-]	[-]	TBD	TBD	1	Near Real Time	Overall quality indicator.
Number of Iterations	[-]	[-]	$0 - n_{iter}^{1dvar}$	1	1	Near Real Time	Number of iterations.
OE Convergence	[-]	[-]	0-1		1	Near Real Time	Convergence flag.

**Figure 5.8:** Extract from IRS Level 2 Atmospheric State Product element list (continued).

(by Grid) and enriched deriving additional information about the potential instability (by Thermodynamics). Such aggregation or “data fusion” step<sup>3</sup> provides a powerful tool to perform, at least in principle, all kind of geospatial (GIS) applications. Within ItDP, this core procedure has been customised for data filtering (looking at the best feature extraction or, equivalently, the best signal-to-noise ratio) and for multivariate analysis to identify and select the predictors. Once again, at source-code level, the use of data structures (as the “dictionary” built-in type in *python*) make easier the categorisation of the classes and the association with the respective populations.

### Data Output

As for Data Input, this procedure is responsible of creating output files at the intermediate and final steps of data processing. From the bunch of data stored inside the processors as internal variables, in theory all the input formats are available for output too, but the greatest advantage is offered by those formats natively “object-oriented”. The reason is that hierarchical data format (as NetCDF or JSON) easily handle associative structures. Considering this, the translation of the internal structures – thought as instances of their respective Data Models – is really straightforward. As an example, the atmospheric

<sup>3</sup>Data fusion is the process of integrating multiple data sources to produce more consistent, accurate, and useful information than that provided by any individual data source.

```

# EXTRACT OBJECT from data
myobj = []
nobj = 0

for i in range(nmask)[-4:]: # only last 4 ranges are considered (i.e. CAPE >= 2500 J/kg)
    if mask[i,:].any():
        for j in range(mask[i,:].sum()):
            x, y = (lon[mask[i,:]][j], lat[mask[i,:]][j])
            print("CLASSE %d - trovato punto #%d (lat,lon): %f %f" % (i,j,y,x))

            # create the MetObj instance
            newobj = MetObj(version='1.0', what='matter') # object attributes
            newobj.add_qty({'centroid_lat':y, 'centroid_longitude':x}) # centroid lat/lon coordinates
            newobj.add_qty({'proc_time':datetime.now().strftime('%Y%m%d%H%M')}) # actual processing datetime (=now)
            newobj.add_qty({'timestamp':timestamp}) # TEMPORARY, WAITING FOR ACTUAL TIMESTAMP of single FOVS!!!
            newobj.add_qty({'class_of_instability':i}) # instability class

            myobj.append(newobj)

            nobj += 1

# EXPORT OBJECT, if any, to json file (serializing the list of objects)
if nobj > 0:
    fout = open(fileout, 'w')
    json.dump(myobj, fout, default=obj_dict)
    fout.close()

```

Figure 5.9: Extract of source code from the Object Detection procedure.

instability objects identified by the Object Detection procedure are encoded in JSON format as output.

## Statistics

The Statistics procedure provides the analytical tools to infer the relationships across the data. The main objective is, within ItDP, to get the correlations of the predictors (prepared by the Classifier procedure) with the events (the weather phenomena as strong convection). Classical regression or fitting algorithms are also available for such correlation analysis. State-of-art external scientific libraries (as *scipy*, see after) have been interfaced with this tool to ensure the respect of a standard and rigorous mathematical treatment.

## Object Detection

This core procedure is designed to complement the Classifier, performing the feature extraction applying the canonical methods of Machine Learning in the field of supervised or unsupervised object detection. An example of functionality implemented here is the clustering module, through the DBSCAN algorithm, used in the ItDP toolkit to identify meteorological structures responsible of strong convection (see figure 6.16). A further added value is the evaluation of the degree of severity of the weather objects found, choosing the proper scores obtained as by-product of the clustering method. Shape, area or density of the objects are of course good indicators. In figure 5.9 is reported a *python* code snippet, extracted from this procedure, that creates the (meteorological) object instance defining a set of key attributes.

### Quality Control

The Quality Control (QC) procedure provides the tools to flag the data with associated quality indicators, to be considered by the other processing modules as highlighted before. An example is the preprocessor that reads the IRS proxy metadata to perform the masking of input data, rejecting the incoming poor quality atmospheric profiles. Similarly, at the end of processing chain the quality indicators of output data are assigned by the QC procedure.

### Map

The Map procedure is specifically designed for mapping purposes, embedding all the modules related to the geographical mapping, transformations (as reprojections) and operators. It provides also the encoding routines to write the output products. The standard graphical formats (as PNG, PDF) as well as the modern GIS-oriented formats (as GML, GeoJSON, KML), specifically addressed to the object classification, are available and ready to use.

### Plot

This procedure performs all kind of plotting of cartesian diagrams (as scatter plot, distribution histograms), being interfaced with the powerful libraries provided by *matplotlib*. In parallel, plotting of thermodynamic diagrams (Skew-T) has been customised and implemented for ItDP toolkit.

---

## 5.2.2 Supervisor System

---

### Data Switching

The Data Switching procedure controls and oversees all the data transfers within the processors. It performs data download, upload and synchronisation. Internal (Local Area Network, LAN) and external (Wide Area Network, WAN) connections with data servers are managed. The *linux shell scripting* language is used as back-end.

### Job Scheduler

The Job Scheduler is designed as the central facility to trigger and manage the procedures of ItDP envisaged for an automatic execution. To accomplish this objective, it provides the mandatory functionalities as the scheduling of the tasks and the logging (execution reports) of the procedures. Furtherly, it offers the possibility of queueing the processes, a particular feature really beneficial in case of system recovery or software changes. Finally, it makes

available some “sleeping tasks” that can be activated in parallel with the online NRT processor in case of reprocessing (offline) experiments. As introduced in section 4.2, the Job Scheduler engine of ItDP toolkit is based on the free and open-source software *ecFlow* developed and maintained by ECMWF<sup>4</sup>, widely used by the community of Meteorological Services.

### System Maintenance

This procedure deals with all the tasks operating at machine level and responsible of cleaning, archiving, backup and recovery of the data. Besides this, it continuously monitors the health of the whole system, reporting any fault or critical issue in the system log, and keeps record of the machine workload for the evaluation of performance bottlenecks.

### GUI

The Graphical User Interface (GUI) is the front-end procedure of both Job Scheduler and System Maintenance, allowing the users to oversee the entire software operational chain. For ItDP the tool provided by the GUI of *ecFlow*, named *ecFlowUI* has been implemented. A snapshot of GUI for the ItDP online and offline processors is shown in figure 4.2 at page 38.

---

### 5.2.3 External Libraries

---

The ItDP procedures rely on a large collection of 3rd party system libraries, entirely free and open-source to be in line with the concept of the whole toolkit. The table 5.2 provides a synthetic list of such packages, with their respective web resources for any additional information needed. The native system libraries (as math, graphical, X, etc.) are not included in the list.

---

<sup>4</sup>More details are available on line at <https://confluence.ecmwf.int/display/ECFLOW>.

**Table 5.2:** List of main 3rd party libraries used by ItDP toolkit.

Module	Function	Web resource
scipy	python scientific libs	<a href="http://www.scipy.org">www.scipy.org</a>
numpy	python numerical libs	<a href="http://www.numpy.org">www.numpy.org</a>
scikit-learn	python Machine Learning libs	<a href="https://github.com/scikit-learn/scikit-learn">github.com/scikit-learn/scikit-learn</a>
matplotlib	python plotting libs	<a href="http://matplotlib.org">matplotlib.org</a>
cartopy	python mapping libs	<a href="http://scitools.org.uk/cartopy">scitools.org.uk/cartopy</a>
proj	map projection libs	<a href="http://proj4.org">proj4.org</a>
NetCDF	NetCDF decoding/encoding libs	<a href="http://www.unidata.ucar.edu/software/netcdf">www.unidata.ucar.edu/software/netcdf</a>
HDF5	HDF5 decoding/encoding libs	<a href="http://www.hdfgroup.org/solutions/hdf5">www.hdfgroup.org/solutions/hdf5</a>
BUFR	BUFR decoding/encoding libs	<a href="https://confluence.ecmwf.int//display/BUFR">confluence.ecmwf.int//display/BUFR</a>
GRIB	GRIB decoding/encoding libs	<a href="https://confluence.ecmwf.int//display/GRIB">confluence.ecmwf.int//display/GRIB</a>
JSON	JSON decoding/encoding libs	<a href="https://docs.python.org/2/library/json.html">docs.python.org/2/library/json.html</a>

### 5.3 An example of end-to-end processor

The procedures described in the previous section are, to some extent, the building blocks of the current ItDP toolkit. At the same time, their modularity and interoperability features ensure that they can be prefigured as the backbone of upcoming software enhancements in support of future research activities and operational applications on nowcasting. To demonstrate how the procedures interact each other to build the ItDP applications needed for the IRS demonstration studies, an example of end-to-end processor is presented here.

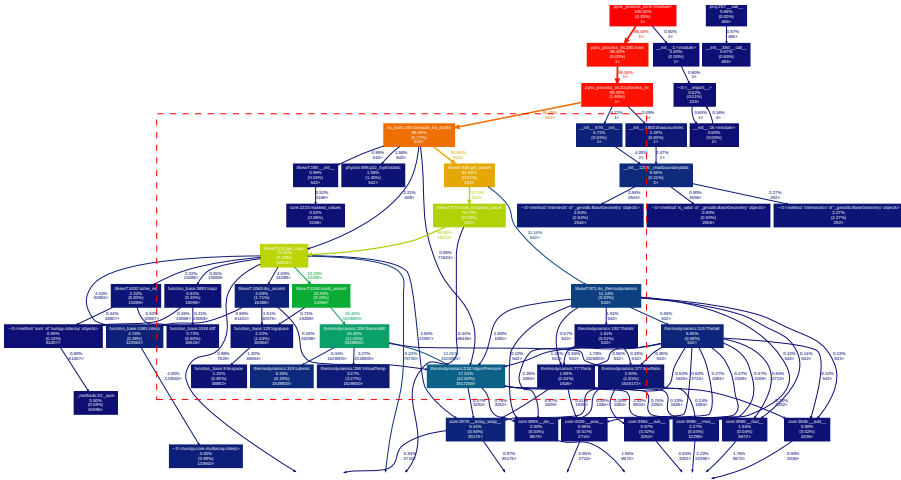
#### Function

To evaluate the signal of potential convective instability carried by a granule of IRS proxy data, data downloaded from EUMETSAT flow through a set of ItDP cascading processors, that decode and preprocess the input data (Data Input, Quality Control), colocating observation and auxiliary information in space and time (Grid). The derived instability indices (Thermodynamics) are then used as predictors (Classifier) for the subsequent correlation analysis (Statistics). In parallel a diagnostics looking at the eventual meteorological clusters is performed (Object Detection). All the output information is stored for further needs (Data Output), including pictorial maps and geospatial data (Map) and the relevant statistical plots (Plot).

#### Source Code

The processing modules cited above are interfaced at *python* scripting level,



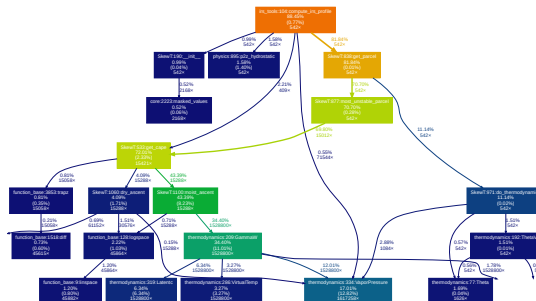


**Figure 5.10:** Extract from the code profiler output for IRS end-to-end processor implemented in ItDP. Connections indicate the relationships across the software modules. Hierarchical structure (top-bottom) and execution times (percentage of total time and number of calls) of the individual modules are shown.

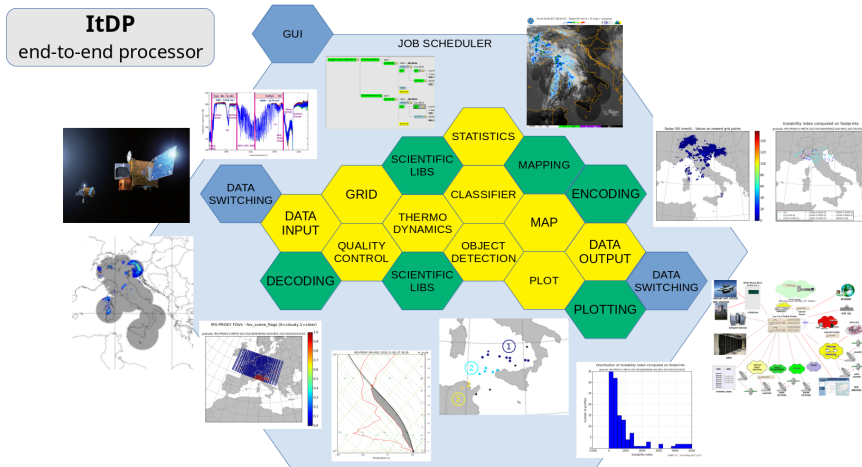
creating a main processor that sequentially calls them. To provide an example of such software interactions, some intuitive plots generated by the code execution profiling may help. In figures 5.10 and 5.11 the relationships and execution times of the individual modules are shown.

### Scheme

The whole software chain is nested into the Supervisor Procedures, that guarantee a resilient management of the system and software processes and enhance the reliability and timely execution of the applications. The functional scheme of this end-to-end processor is naively sketched in figure 5.12.



**Figure 5.11:** Extract from the code profiler output for IRS end-to-end processor implemented in ItDP (detail).



**Figure 5.12:** Example of end-to-end processor, showing the main functional blocks identified as ItDP procedures. Supervisor (blue) and processing (yellow) blocks, as well as external libraries (green) have been interconnected to realise a stand-alone, cascading processor for IRS proxy and auxiliary data.

## 5.4 What's next?

The software modules designed, tested and implemented in the operational processors of IRS proxy data represent the first “release” of the ItDP toolkit, intended as contribution to the demonstration studies on MTG-IRS and as a “basket of applications” specifically addressed to support nowcasting and very short range forecast.

The procedures illustrated so far may be developed to gather new features in such research domain. The inclusion of new kind of data, of new algorithms of machine learning or new processing chains for data handling are all straightforward activities, at least in principle considering the architectural design of ItDP.

Referring to the Computer Science aspects, it is worthwhile to mention that an interesting field of investigation regards the use of virtualised environments, especially in the perspective of a large production of data. The virtualisation of the hardware, software and middleware layers, indeed, allow the best dynamic allocation of the resources, adjusted to the expected or actual system workload. By the way, the ItDP toolkit has been implemented at COMET on a state-of-art virtual infrastructure.

---

A further option to be evaluated for future developments is related to the use of *docker*<sup>5</sup>, a tool offering additional advantages with respect to the virtual machines, in particular for scalable applications or software run through parallel instances with different configurations and release versions. Just to note, EUMETSAT L2VDP service is based on the utilisation of the *docker* framework.

---

<sup>5</sup>Docker is a software application that performs operating-system-level virtualisation, also known as “containerization”. Containers are isolated from each other and bundle their own application, tools, libraries and configuration files; they can communicate with each other through well-defined channels. All containers are run by a single operating-system kernel and are thus more lightweight than virtual machines.



# 6

## CASE STUDIES

---

*This chapter provides an overview of the significant case studies collected within ItDP project and selected to highlight some key features of IRS proxy data and their potential impact on nowcasting applications.*

### Contents

---

6.1	Colocation and analysis of the problem . . . . .	65
6.2	Revisiting time . . . . .	70
6.3	Object detection . . . . .	73
6.4	Impact of a priori state . . . . .	76

---

During the deployment of the EUMETSAT IRS NRT Demo Project a large volume of proxy data were available at COMET both in near real time mode and in offline mode as reprocessing experiments. Lasting about one year from mid-2016, a valuable amount of information was analysed to explore the potential applications in the field of nowcasting on MTG-IRS day-1 operations as well as, from a theoretical point of view, the necessary adaptations and the further improvements of the L2 Processor. To investigate the correlation between IRS proxy data and actual strong convective events, a list of candidate case studies has been prepared, also relying on the automatic selection criteria configured for the ItDP online processor and based on geographical bounding box, CAPE and radar reflectivity thresholds, quality indicators of satellite data, etc. Regarding the area of interest, the Italian domain has been preferred due to the availability of high density observations, such as lightnings, rain-gauges, weather radars, specifically designed to resolve the typical temporal and spatial scales of convection. Concerning the thresholds for instability indices and for observations as radar reflectivity, the standard values taken from literature specific for Euro-Mediterranean region have been chosen (see for example references [12] and [13]). By the way the a posteriori analysis of the events confirmed the good agreement with this choice. Being the data delivery service offered by EUMETSAT L2VDP almost uninterrupted for the whole 1-year period of the project – even though not considered strictly as an operational service by EUMETSAT but on the best effort basis – and thanks to the high reliability of ItDP online processor in order to process ALL the incoming data, a good number of “tagged” events has been identified and recorded. An extract of this list is shortly reported in figure 6.1.

The following sections are intended to examine, one at the time, some key

SEVERE CONVECTION CASE STUDIES (since 2015)			
#	when	where	notes
...	...	...	...
20	01 Feb 2015	Greece	<a href="http://emergency.copernicus.eu/mapping/list-of-composites/EMSR117">http://emergency.copernicus.eu/mapping/list-of-composites/EMSR117</a> <a href="http://www.eumetsat.int/webste/home/image-storage/Library/DAT_2546692.html">http://www.eumetsat.int/webste/home/image-storage/Library/DAT_2546692.html</a>
21	02 Feb 2015	Spain	...
22	05 Mar 2015	Central Italy	<a href="http://emergency.copernicus.eu/mapping/list-of-composites/EMSR118">http://emergency.copernicus.eu/mapping/list-of-composites/EMSR118</a> <a href="http://www.eumetsat.int/webste/home/image-storage/Library/DAT_2570283.html">http://www.eumetsat.int/webste/home/image-storage/Library/DAT_2570283.html</a>
23	15 Apr 2015	Portugal / Spain	thunderstorms
24	15 May 2015	Northern Italy	Milano Malpensa thunderstorms
25	18-19 May 2015	South/Northern Italy	thunderstorms on Po Valley, severe weather in Salento
26	21 May 2015	Northern Italy	thunderstorms on Milano, Lombardia
27	27 May 2015	Central Italy	thunderstorms with hail on Rome, Lazio
28	8 July 2015	North-Eastern Italy	thunderstorms / super-cell and funnel cloud on Venice area, Veneto <a href="http://fonti.redeveneto.corriere.it/venezia/tempeste/tempeste/2015/07/08/milano-tempo-tomba-d-aria-mira-2301627592345.shtml">http://fonti.redeveneto.corriere.it/venezia/tempeste/tempeste/2015/07/08/milano-tempo-tomba-d-aria-mira-2301627592345.shtml</a> <a href="http://video.corriere.it/08-07-2015/tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/">http://video.corriere.it/08-07-2015/tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/</a> <a href="http://www.rai.it/rad5/080715">http://www.rai.it/rad5/080715</a> <a href="http://www.meteo.it/komali08-luglio-2015-forse-tornado-su-venezia-la-testimonianza-7490.shtml">http://www.meteo.it/komali08-luglio-2015-forse-tornado-su-venezia-la-testimonianza-7490.shtml</a>
29	29 July 2015	North-Eastern Italy	thunderstorms on Eastern Lombardia and Veneto <a href="http://www.emeteeo.org/2015/07/29/milano-on-lombardia-forse-temporale-su-bergamio-E-video-della-tempesta-da-almemno-san-savatore/">http://www.emeteeo.org/2015/07/29/milano-on-lombardia-forse-temporale-su-bergamio-E-video-della-tempesta-da-almemno-san-savatore/</a> <a href="https://www.youtube.com/watch?v=i6LBJ3PFlag">https://www.youtube.com/watch?v=i6LBJ3PFlag</a>
30	1 August 2015	Central Italy	super-cell thunderstorms on Florence area, Tuscany <a href="http://www.emeteeo.org/2015/08/01/milano-tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/">http://www.emeteeo.org/2015/08/01/milano-tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/</a>
31	11 August 2015	Southern Italy	flashfloods / thunderstorms on Calabria <a href="http://www.eumetsat.int/webste/home/image-storage/Library/DAT_2738277.html">http://www.eumetsat.int/webste/home/image-storage/Library/DAT_2738277.html</a>
32	1 October 2015	Central/Southern Italy	flashfloods / thunderstorms on Sardinia and Sicily <a href="http://www.emeteeo.org/2015/10/01/milano-tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/">http://www.emeteeo.org/2015/10/01/milano-tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/</a>
32	5 August 2016	Northern Italy	flashfloods / thunderstorms on Milano area, Lombardia and Treviso area, Veneto <a href="http://milano.corriere.it/2016/08/05/tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/">http://milano.corriere.it/2016/08/05/tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/</a> <a href="http://fonti.redeveneto.corriere.it/venezia/tempeste/tempeste/2016/08/05/milano-tempo-tomba-d-aria-mira-2301627592345.shtml">http://fonti.redeveneto.corriere.it/venezia/tempeste/tempeste/2016/08/05/milano-tempo-tomba-d-aria-mira-2301627592345.shtml</a> <a href="http://ra24.liv.413245/condizioni/2016/08/05/milano-tempo-tomba-d-aria-mira-2301627592345.shtml">http://ra24.liv.413245/condizioni/2016/08/05/milano-tempo-tomba-d-aria-mira-2301627592345.shtml</a>
33	6 Nov 2016	Central Italy	tornado-like thunderstorm on Lazio coastal area <a href="http://roma.corriere.it/2016/11/06/tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/">http://roma.corriere.it/2016/11/06/tempeste-veneto-25ag-11-ec-85c7-ec5c793b38f/</a>
...	...	...	...
...	...	...	...

Figure 6.1: Extract from list of significant convection case studies.

aspects of the matter. It is worth to point out that, since the baseline requirement for day-1 is to deliver to end users the vertical soundings retrieved only on clear sky FOVs<sup>1</sup>, the majority of the meteorological situations here examined refer to the spring-summer season period. This is also related to the climatology of convective events evolving by clear sky conditions with high latent instability, which occur in Euro-Mediterranean area mainly in the same period, including early autumn season as well.

## 6.1 Colocation and analysis of the problem

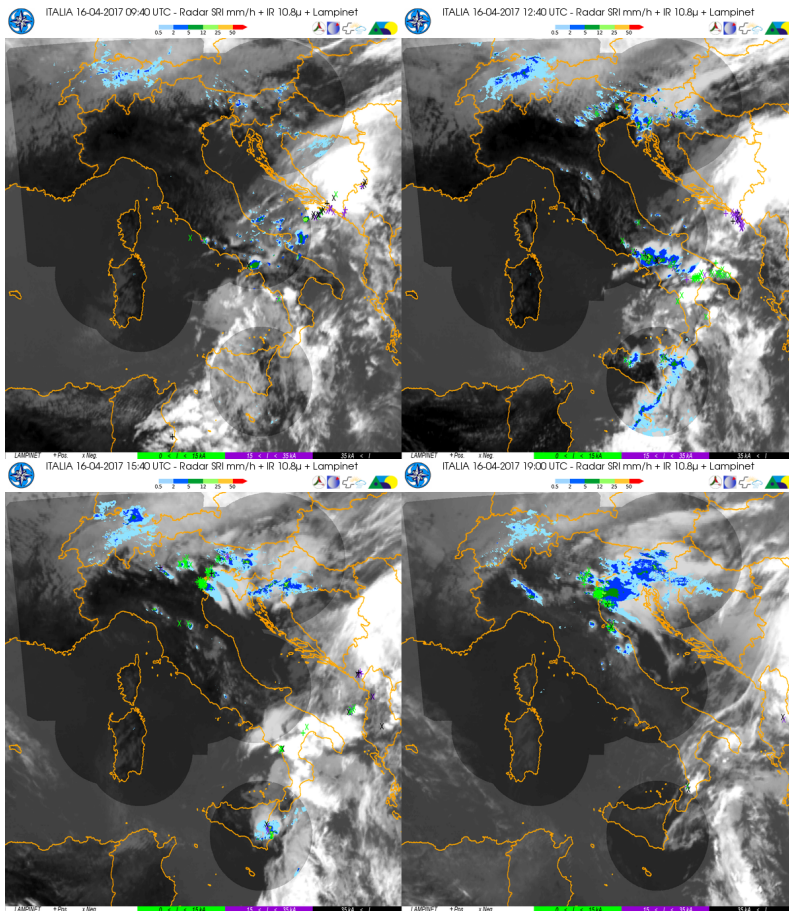
At first glance, the comparison between satellite IR hyperspectral retrievals (named also “pseudo-soundings”) and actual atmospheric thermodynamics may be considered not a complicate exercise, looking for example at the Skew-T diagrams shown in the sections 4.3 and 4.4, very similar to the ones routinely plotted and examined by weather forecasters<sup>2</sup>. However, real world looks different. The two classes of analogue atmospheric variables (e.g. space-retrieved versus actual temperature) have different statistical behaviour, because whereas the latter may be referred as the “true” state, the former is the result of a series of convolutions with many sources of uncertainty. In particular, it is intrinsically correlated with the “a priori” state of the atmosphere, necessary for the inversion algorithm but potentially far from the true state (for theoretical aspects of the retrieval principles, see [10]). In other words, what looks similar not always is similar in statistical sense. This statistical difference results in potentially different spatial scales of features or, more properly, in different covariance matrices. To give a flavour of evidence of such pure theoretical considerations and to show the complex nature of the problem of signal extraction from space-based observations, the following case study is worth to be analysed.

On 16 April 2017 evening, a line of thunderstorm cells affected north-western Italy, as shown at 19.00 UTC in the composite image (IR satellite-radar-lightning data) sequence shown in figure 6.2. The first image in the sequence refers to 09.40 UTC situation, exactly the time of overpass of IASI-B<sup>3</sup> over northern

<sup>1</sup>The same requirement was applied by EUMETSAT to IRS NRT Demo Project regarding L2 data delivery.

<sup>2</sup>Just to notice, the wind is not available for pseudo-soundings, because this variable is not included in the atmospheric state of IR 1-dimensional retrieval schemes. This lack of information is crucial, because the vertical wind shear, for example, is strongly correlated to the triggering of severe thunderstorm events.

<sup>3</sup>IASI on board METOP-B.

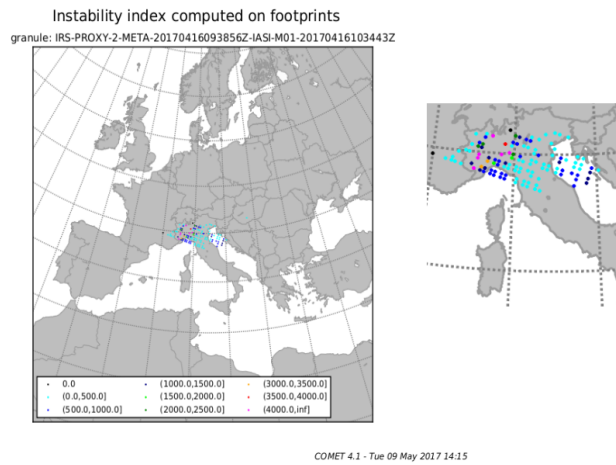


**Figure 6.2:** Sequence of composite images (IR satellite-radar-lightning data) for 16 Apr. 2017 case study. Top-left: obs+0h, top-right: obs+3h, bottom-left: obs+6h, bottom-right: obs+9.3h.

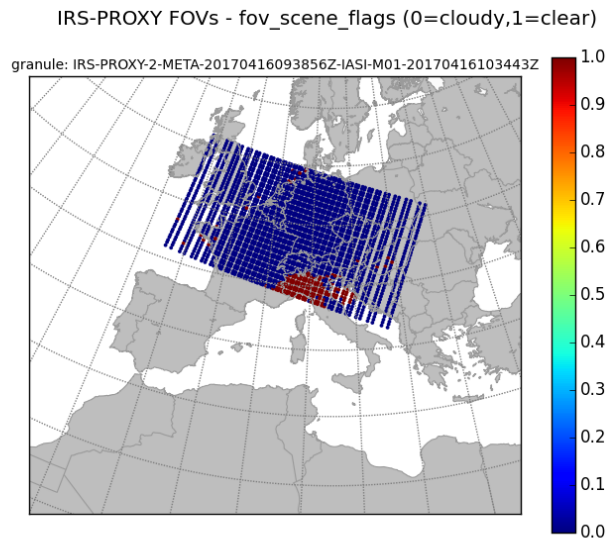
Italy, with clear sky conditions on north-west, low level clouds on north-east and overcast with embedded thunderstorm cells in central and southern Italy.

The map of CAPE index retrieved from IRS proxy data generated from the coincident IASI granule at 09.38 UTC is shown in figure 6.3. High values of CAPE are associated to some FOVs over north-western area, lower values on the rest of area. It is important to remind that retrieval is available only for clear sky FOVs, therefore the plot of the scene cloud flag, included in the data granule, must be considered (figure 6.4). The cloudy conditions on north-east are here evident, in line with IR imagery from SEVIRI shown before, demonstrating that IRS proxy cloud classifier criteria are efficient.

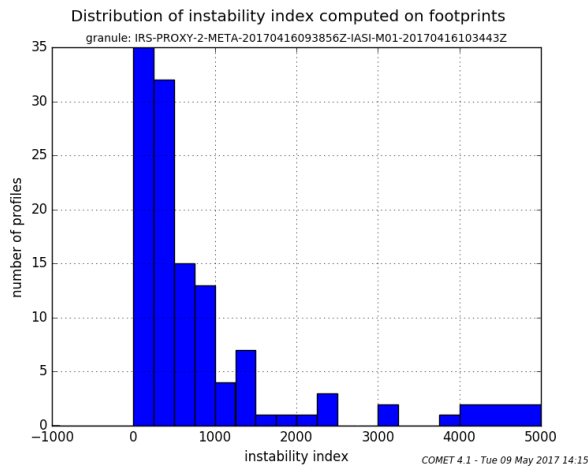




**Figure 6.3:** Map of IRS-proxy CAPE instability index for 16 Apr. 2017 case study.



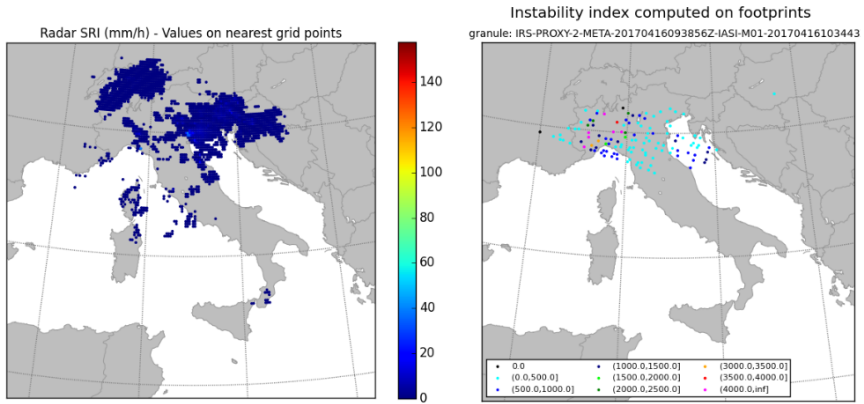
**Figure 6.4:** Map of IRS-proxy cloud mask for 16 Apr. 2017 case study.



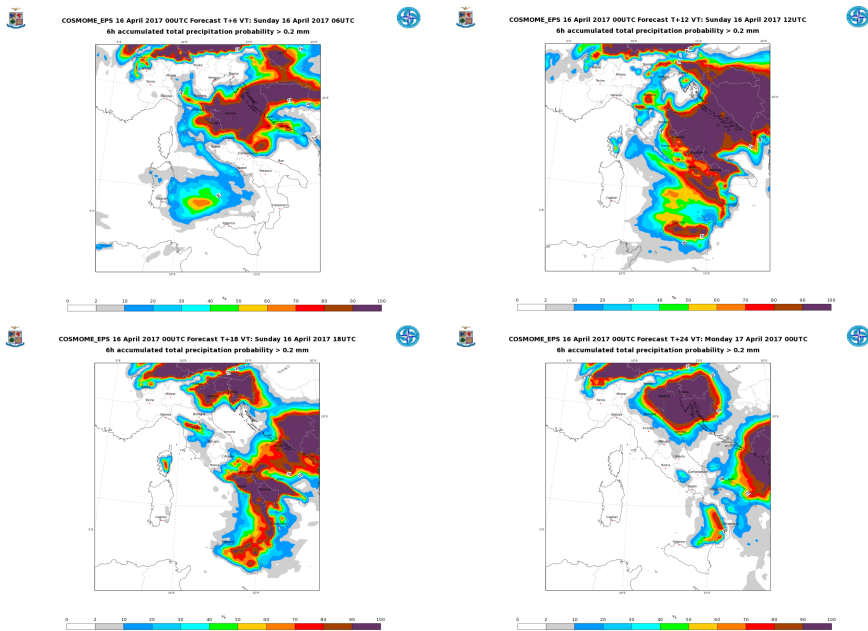
**Figure 6.5:** Distribution of IRS-proxy CAPE instability index for 16 Apr. 2017 case study.

The presence of some high valued CAPE FOVs could be further analysed in the distribution histogram, showing such outliers on the right of figure 6.5. Associating the “signal” to the CAPE at IASI timestamp (obs + 0h) and the “event” to the radar reflectivity field at the time of identified event (obs +9.3h), the next step is to perform a geographical colocation of the two fields, mapping both to a common grid, as demonstrated in figure 6.6. This step is crucial for the correlation analysis and, in general, for feature extraction applications as for data mining. Thus the choice of the interpolation grid and the algorithm for the interpolation (in this case the nearest point value with an effective radius filter) must be carefully evaluated. Of course, cloudy grid-points may be neglected in further analysis for example by masking the grid.

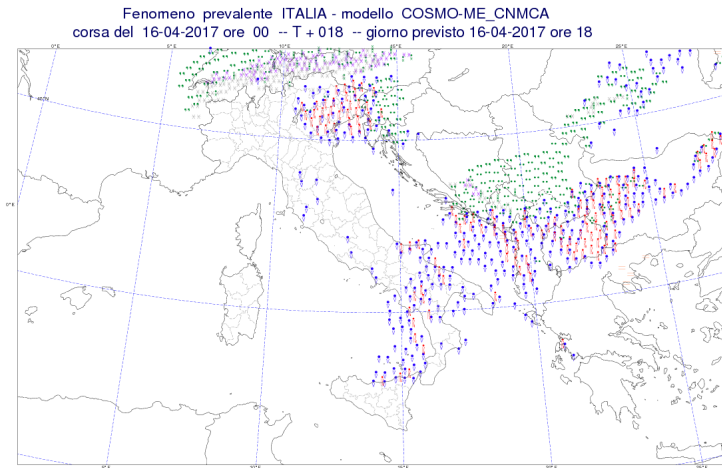
Referring to the detection of the signal over north-west, this case study shows the good correlation of convective events with CAPE from IRS pseudo-soundings more of 9 hours ahead, a very appealing early time for nowcasting purposes. Furtherly, the forecast provided by NWP model for this case totally missed the event on north-west, as shown in the plots of COSMO ensemble-model output for probability of precipitation (sequence of 6 hours-step forecasts in figure 6.7). However, the NWP model predicted with good accuracy the precipitation events on north-eastern Italy (see also figure 6.8), missed by the analysis of retrieved CAPE, due to the missing cloudy FOVs and, probably, to partly cloudy contaminated FOVs.



**Figure 6.6:** Spatial-temporal colocation of IRS-proxy data and observed event for 16 Apr. 2017 case study. Left: event at 19.00 UTC, right: observation at 09.40 UTC.



**Figure 6.7:** Sequence of 6 hours-step probability forecasts of COSMO model for 16 Apr. 2017 case study.



**Figure 6.8:** Significant weather map (post-processing product) from deterministic forecast of COSMO model for 16 Apr. 2017 case study.

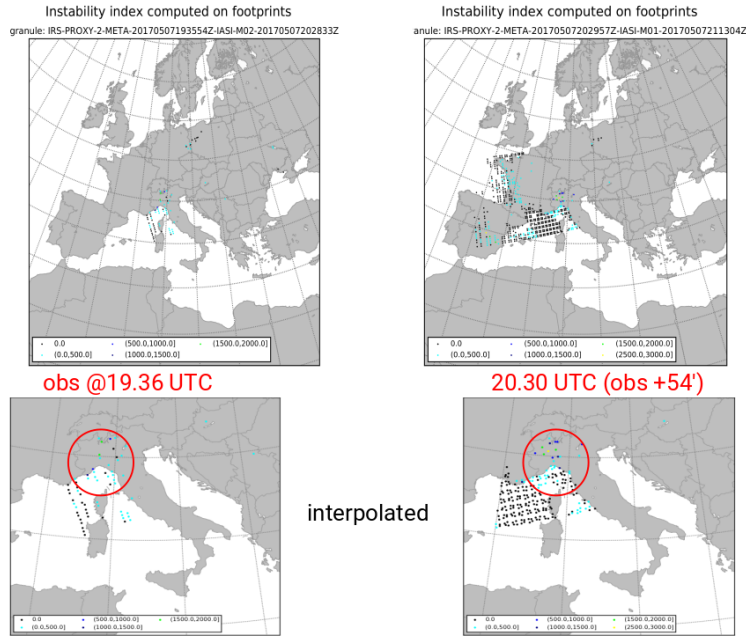
To summarise, this case study demonstrates the potential feature extraction of signal by collocation of IRS proxy observations (as “predictor”) with the events (as “truth”). Particular care must be taken regarding the thresholds of predictors, the interpolation algorithm and the scene flags (cloudy or partly cloudy contaminated FOVs).

## 6.2 Revisiting time

A very appealing feature offered by the Metop-A/B orbital phase shift (50 minutes of delay on the same orbit<sup>4</sup>) is the possibility to take advantage of a short “revisiting time” (less than 1 hour) to “reinforce” – borrowing this word from the reinforcement learning methods – the observed signature of a severe weather event, as demonstrated in a couple of case studies hereafter.

In figure 6.9 are plotted the maps of predictor (CAPE from pseudo-soundings) on both original and interpolated grids for METOP-A (obs at reference time 7 May 2017, 19.36 UTC) and subsequently for METOP-B at 20.30 UTC (obs+54’).

<sup>4</sup>With new METOP-C, a third frequent overpass will be possible for points along satellite tracks.



**Figure 6.9:** Example 1 of revisiting time for 7-8 May 2017 case study – proxy data.

The common interpolation grid shows an overlap region on north-western Italy (circled in red) with moderately high values of CAPE.

On 8 May, at 5.00 UTC (obs+9h) a cluster of thunderstorm rapidly grew up exactly on this region, as evident from composite (IR satellite-radar-lightning data) image in figure 6.10.

The same day, 8 May, after a pause of convective activity (allowing clear sky conditions) in the late evening new thunderstorm cells developed in the same area. The overpasses of METOP-B firstly (obs at reference time, 20.09 UTC) and METOP-A secondly at 20.54 UTC (obs +45') flagged an overlap area with high values of CAPE (figure 6.11), where convective events occurred at 21.40 UTC (obs+90'), as shown in figure 6.12.

This case studies demonstrate, especially in the perspective of the geostationary orbit of MTG-IRS, that frequent hyperspectral soundings are able to catch a signal persisting in the atmosphere on convective time scales (hourly to sub-hourly) and reinforced by two independent measurements. Modern algorithms based on reinforcement learning are appealing to exploit such characteristics. To this regards, it is worth to point out that even nowadays the opportunity given by the METOP satellites constellation is under-exploited, as the majority

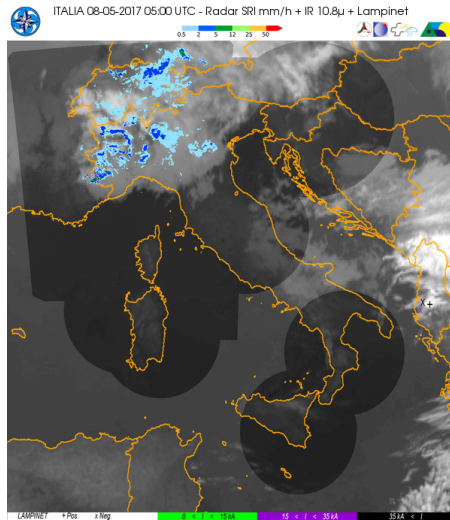


Figure 6.10: Example 1 of revisiting time for 7-8 May 2017 case study – event.

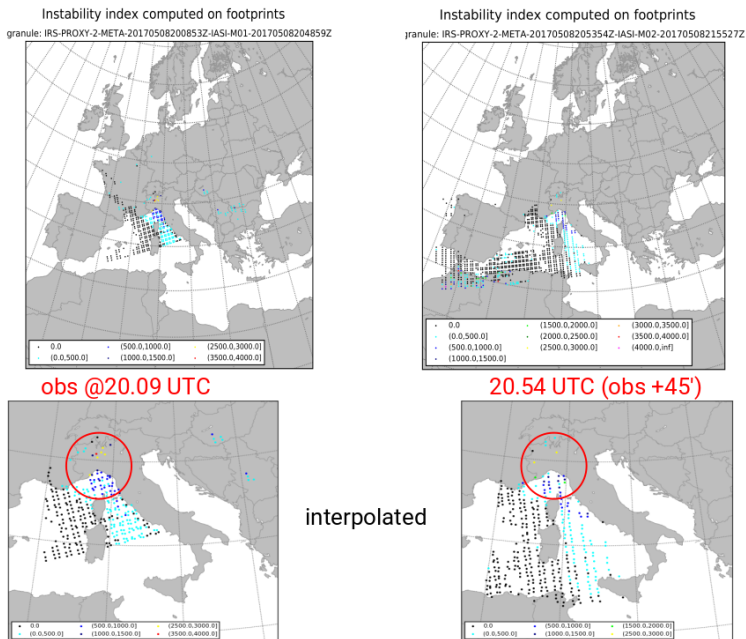
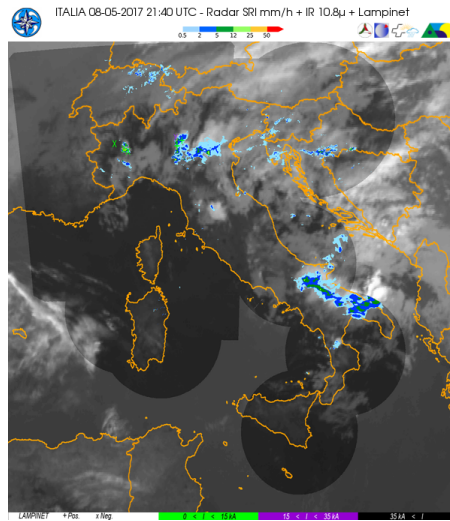


Figure 6.11: Example 2 of revisiting time for 7-8 May 2017 case study – proxy data.



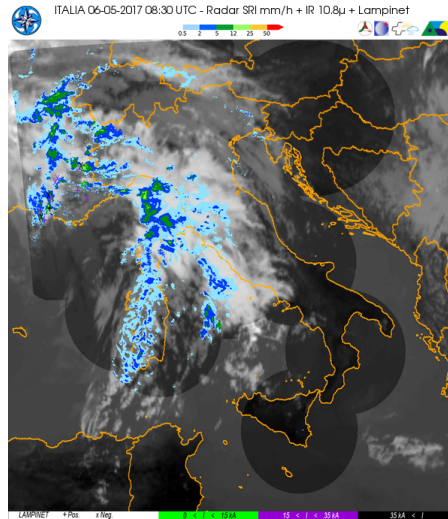
**Figure 6.12:** Example 2 of revisiting time for 7-8 May 2017 case study – event.

of current detection and tracking algorithms of nowcasting applications usually neglect reinforcement learning mechanisms and correlation of features in the time domain, apart simple advection schemes.

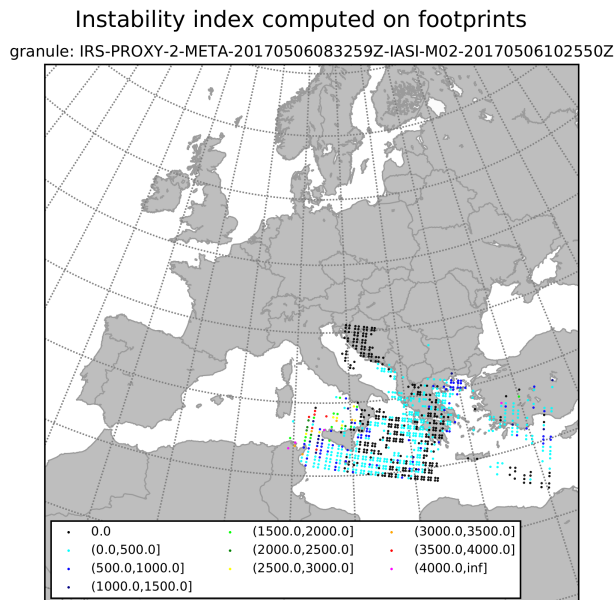
## 6.3 Object detection

The next case study is intended to give an example of the potential applications on nowcasting of innovative concepts involving Machine Learning algorithms, with particular emphasis on spatial clustering methods. On 6 May 2017 morning, Italy was interested by an intense south-western warm advection, as evident from composite image at 08.30 UTC shown in figure 6.13.

The map of CAPE predictors computed from METOP-A overpass almost at the same time (08.32 UTC) shows, on clear sky FOVs on the south-eastern edge of the convergence line over Tirrenian Sea, a set of high valued points (figure 6.14). These point are apparently randomly distributed in space, with neither privileged directions nor evident groups, as shown in the zoom in figure 6.15.



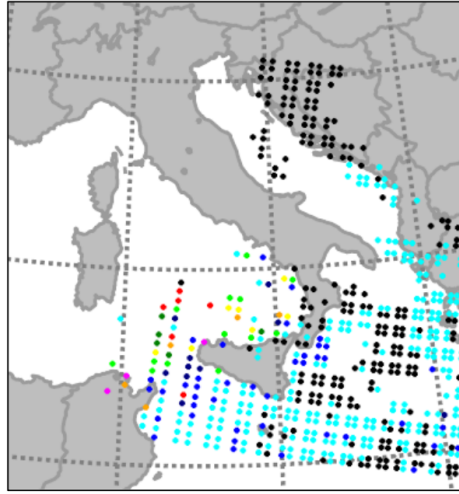
**Figure 6.13:** Composite image (IR satellite-radar-lightning data) for 6 May 2017 case study.



COMET 4.1 - Sat 06 May 2017 10:50

**Figure 6.14:** Map of IRS-proxy CAPE instability index for 6 May 2017 case study.



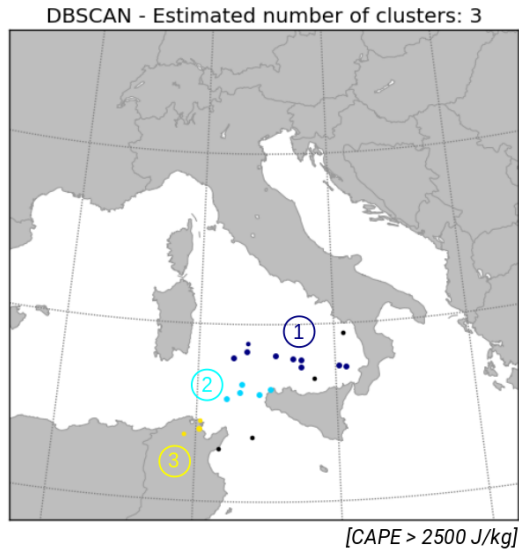


**Figure 6.15:** Zoom of the map of IRS-proxy CAPE instability index for 6 May 2017 case study.

Applying the clustering algorithm DBSCAN<sup>5</sup> (Density-Based Spatial Clustering of Applications with Noise) to this set of points, filtered by a moderate lower threshold on CAPE values (2500 J/kg), the result is a bit surprising, because the unsupervised algorithm is able to automatically detect three clusters, as shown in figure 6.16. The interesting feature, perfectly recognised by the clustering method, is the identification of three separate group of points, associated only by spatial proximity and density, which are aligned in the south-west to north-east direction. Translating it to the physical meaning of the variables involved in the problem, the algorithm detects the existence of three “patterns” of potential unstable air-mass along the main stream of advection, and more precisely on the eastern edge (warm sector). In nowcasting sense, the advection of unstable air-mass by successive pulses is taking place, according to the classical conceptual model for this family of Mediterranean convective system.

---

<sup>5</sup>The DBSCAN algorithm is an unsupervised clustering method that views clusters as areas of high density separated by areas of low density. Due to this rather generic view, clusters found by DBSCAN can be any shape, as opposed to k-means method which assumes that clusters are convex shaped. The central component to the DBSCAN is the concept of *core samples*, which are samples that are in areas of high density. The outliers are filtered out introducing a control parameter that refers to the noise variance. For more details on DBSCAN implementation, see the python *scikit-learn* package documentation available on line at <https://github.com/scikit-learn/scikit-learn>.



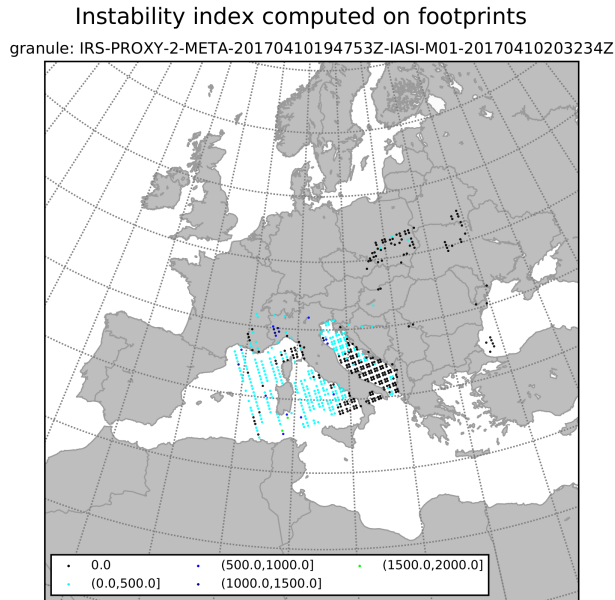
**Figure 6.16:** Example of object detection by spatial clustering for 6 May 2017 case study.

This case study demonstrate that the use of spatial clustering methods may be beneficial to extract the information content from dense observed data – as high-resolution satellite data – related to the length scales of severe weather patterns, detecting automatically the meteorological “objects”.

## 6.4 Impact of a priori state

As stated in the section 6.1, the statistical nature of the problem, arising from the algorithm of extraction of the information content (or signal) from hyperspectral measurements, namely the inversion of spectral radiances through classical retrieval techniques, requires to some extent a knowledge of the a priori state of the atmospheric variables involved. The following case study, based on a reprocessing experiment, demonstrates the sensitivity of the final retrieved soundings to the particular source of a priori information chosen.

On 10 April 2017, convective events affected northern Italy during afternoon

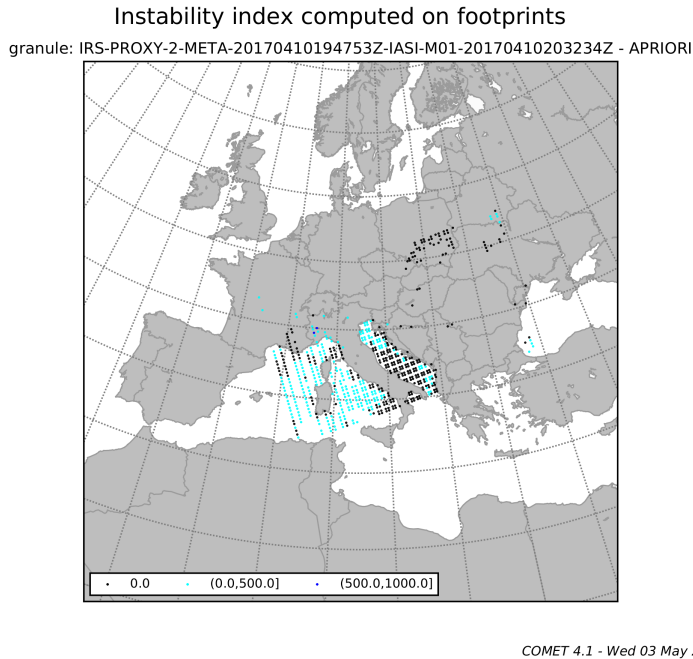


**Figure 6.17:** Map of IRS-proxy CAPE instability index for 10 April 2017 case study - NRT data.

and night time. The associated METOP-B overpass at 19.47 UTC provided the input to the IRS NRT proxy data generation, resulting in derived pseudo-soundings available for ItDP online processor at COMET. The related map of CAPE index computed on such granule, corresponding to a 3' portion of the orbit, is shown in figure 6.17.

It is worth to remark that, according to the requirements of L2VP, proxy data delivered to investigators include, besides the retrieved profiles (corresponding to the so-called “posterior” state in the Bayesian theory), all atmospheric variables of the a priori state. Taking advantage of such opportunity, the representation of the a priori CAPE index over the identical FOVs is straightforward. The result is shown in figure 6.18. It is easy to notice in the map a slight change in the CAPE intensity for some FOVs of north-western Italy, located in the area of interest.

This difference ( $\Delta$ CAPE) may be interpreted, according to the literature studies cited in chapter 3, as a potential signal detection by the cascading processors. From a theoretical point of view, this is justified noting that the retrieval algorithm, as in general any inversion method, is properly defined to produce the “adjustment” of the posterior state variables to the analysis state, which



**Figure 6.18:** Map of IRS-proxy CAPE instability index for 10 April 2017 case study - NRT a priori data.

maximises, in likelihood statistical sense, the correction of the a priori “rough” information exploiting the signal carried by the measurements.

Thanks to the availability, within IRS Demo Project, of reprocessed proxy data for the same granule, a further analysis was possible on this topical aspect. The relevant hypothesis here is that IRS proxy reprocessed data are generated using a different prior state to be ingested by the retrieval scheme. Such a priori information has been widely demonstrated by the MIMAG experiments to have better quality, because it is derived from a more accurate and reliable atmospheric estimate at the right time of the observation, based on the Ensemble Data Assimilation (EDA) system of ECMWF Centre, one of the most skilled analysis worldwide.

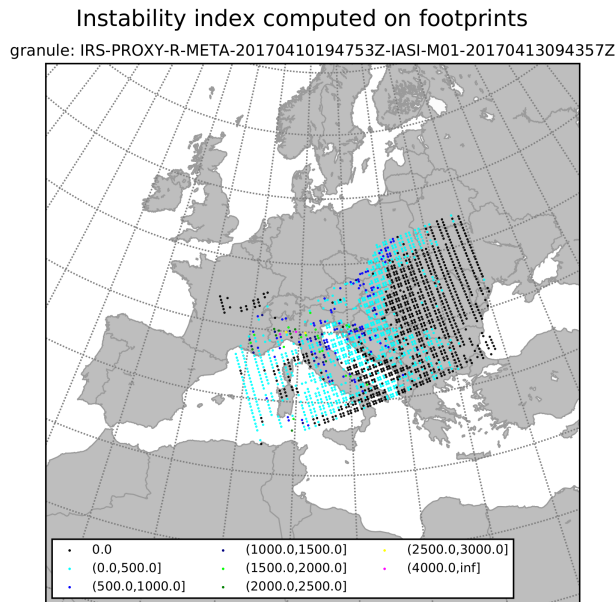
For timeliness constraints, Level 2 proxy data generated in near real time mode cannot use the same a priori as reprocessed data. The reason is that EDA analysis is generally not available at the time of satellite observations, because of a unavoidable lag necessary for the collection over a relatively wide time window (6 hours, at least) of the data coming into the data assimilation system. Therefore IRS NRT proxy data are generated by L2VDP using a short cutoff analysis-forecast system, also provided by ECMWF, but with coarser time

resolution and so potentially far from the true state of atmosphere.

The proof of the better quality of the a priori information for reprocessed data, by one side, and of the ability of  $\Delta$ CAPE to detect the convection signal and its features, is evident looking at the figures 6.19 and 6.20. The difference on the north-western area of Italy between posterior and prior CAPE index is now more striking.

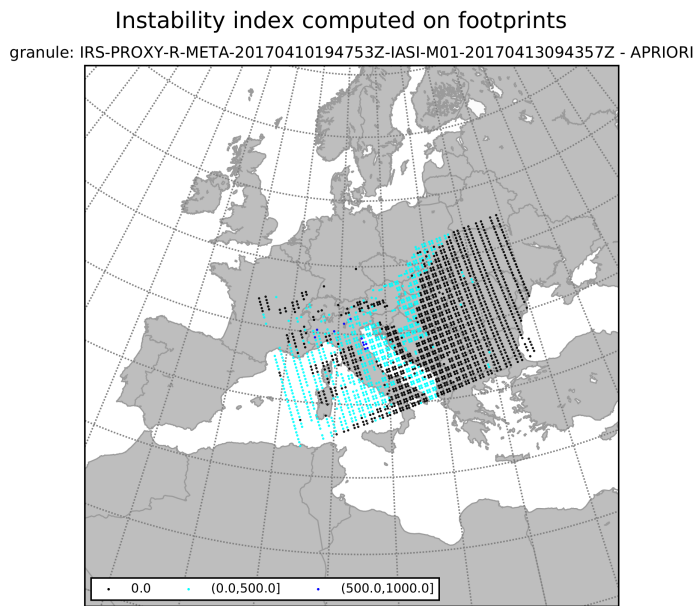
Another implication of the better quality of a prior atmospheric state is the surprisingly larger number of available FOVs of proxy data, comparing as example figures 6.17 and 6.19. This is related, of course, to the greater number of successful retrievals obtained in the case of the better both analysis and collocation in time domain provided by EDA system.

To summarise, this case study demonstrates that the choice of the a priori information to be used for the inversion algorithm has a very strong impact on the quality of the retrieval, which in turn affects the ability of catching the signal of atmospheric instability by any post-processing system. Furtherly, the use of incremental indices (i.e. posterior with respect to prior) for signal detection is really promising as better predictor than absolute indices for posterior state alone.



COMET 4.1 - Tue 02 May 2017 17:19

**Figure 6.19:** Map of IRS-proxy CAPE instability index for 10 April 2017 case study - reprocessed data.



COMET 4.1 - Tue 02 May 2017 17:22

**Figure 6.20:** Map of IRS-proxy CAPE instability index for 10 April 2017 case study - reprocessed a priori data.

# 7

## DISCUSSION AND CONCLUSIONS

---

*This chapter summarises the findings of the investigation conducted in the thesis and provides the concluding remarks, giving an outlook on possible applications and future work.*

### Contents

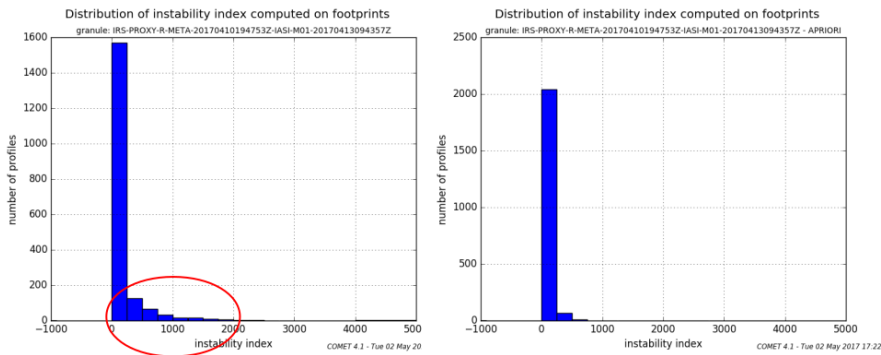
---

7.1	Considerations . . . . .	82
7.2	Applications . . . . .	84
7.3	Further developments . . . . .	85
7.4	Concluding remarks . . . . .	86

---

## 7.1 Considerations

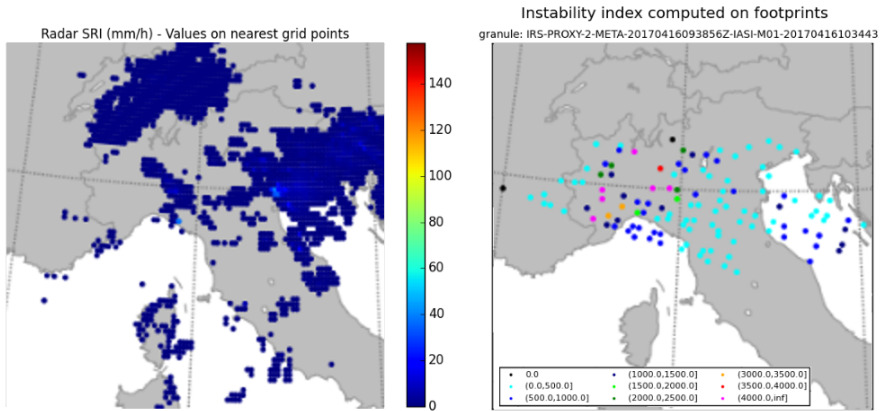
The findings of the previous chapter argue, following an experimental approach, on the hyperspectral signature of the convective instability. In particular, the analysis of the case study in section 6.4 demonstrate that, under an optimal choice of the a priori atmospheric state for the generation by inversion methods of the vertical profiles, the information content carried by the hyperspectral IR sounder observations is correlated to the potential instability of atmosphere, able to trigger convective events or even severe storms. The proof of the existence of such relationship can be inferred looking at the tail of the CAPE distribution in the left histogram plot of figure 7.1, which demonstrates that posterior atmospheric state (retrieved using hyperspectral observation) is able to carry a measurable “instability signal” with respect to the corresponding prior state (using only the first-guess profile from NWP system).



**Figure 7.1:** Comparison of posterior (left) and prior (right) distribution of CAPE for the same observed IRS proxy granule.

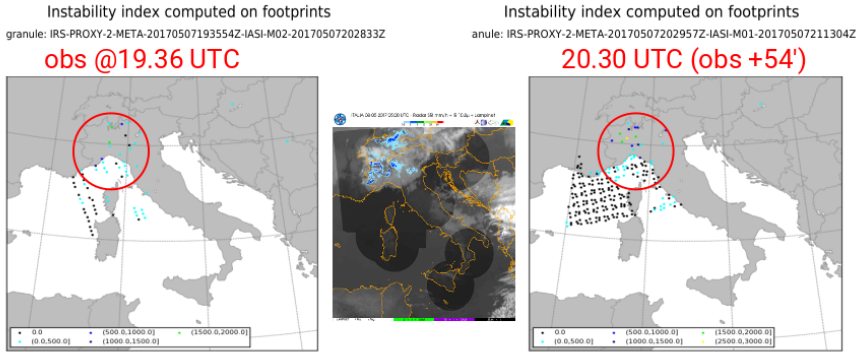
The ability of the hyperspectral sounding of catching the atmospheric latent instability, even well in advance, is a promising feature for nowcasting purposes. As highlighted by a wealth of scientific works (see for example [12]) on the climatology of mid-latitude cyclogenesis (the generation mechanism of storms), both the upper and lower tropospheric thermodynamic profiling in clear sky pre-convective environment has the potential for a satellite signature. As for the 16 April 2017 case study reported in section 6.1, a very good early correlation between instability indices and actual storms can be detected. The figure 7.2 provide a detail on the space-time collocation of the signal (more than 9 hours ahead) and the event.





**Figure 7.2:** Detail of the colocation of IRS-proxy data and observed event for 16 Apr. 2017 case study. Left: event at 19.00 UTC, right: observation at 09.40 UTC.

Another unprecedented feature, looking forward at MTG-IRS, is the higher temporal frequency of the geostationary hyper-soundings, closer to the typical time scales of the convective patterns, with respect to the current polar orbiting satellites revisiting time. A flavour of such peculiarity can be argued thanks to the Metop-series orbital phase shift (50 minutes for Metop A and B). The sub-hourly refresh rate – waiting for the 30 minutes interval provided by IRS – is able to amplify or reinforce the atmospheric signal. The derived time lagged correlation is expected to improve the statistics introduced in the future machine learning systems, by reducing the off-diagonal contributions to the contingency (or confusion) matrix, namely the false alarms and the misses. The figure 7.3 provide a detail on the reinforced information from two successive Metop overpasses and the corresponding storm event for the 7-8 May 2017 case study.



**Figure 7.3:** Detail of revisiting time feature for 7-8 May 2017 case study – proxy data and event.

## 7.2 Applications

The demonstration studies and the software tools made available through this research work may be considered as the starting line for a number of possible applications. From a scientific point of view, the coming calibration and validation activities of IRS Level 1 and Level 2 products to prepare MTG-S day-1 operations may benefit from the know-how and the technology developed here. Further case studies and, possibly, longer experimental NRT proxy data campaigns are expected to be extremely valuable to improve the exploitation of the hyperspectral products.

From the end-user perspective, the online end-to-end processors of NRT data have immediate applications to support the nowcasting and very short range forecasting activities. In particular, considering the timely, synthetic and object-oriented information provided by the ItDP toolkit, the output products are promptly available and ready to use for the forecasters, who can visualise them on their operational workstations.

Finally, the findings of this thesis point out that the pieces of information extracted by the hyperspectral soundings could contribute to the existing GIS platforms operating in the field of automatic or semi-automatic early warning of weather hazards. This kind of data fusion application is envisaged as a breakthrough facilitator for the current and future Spatial Decision Support Systems (SDSS).

---

## 7.3 Further developments

---

An interesting area of investigation for the follow-up studies on this matter is surely related to the object detection, specifically designed for meteorological convective systems. The characterisation of the latter has been shown here as conceptually easy, due to their 3D well defined structures and hyperspectral fingerprints. As demonstrated in section 6.3, the objected oriented approach for the detection and tracking is a potential basin of further research activities. As an example, the Method for Object-Based Diagnostic Evaluation (MODE), applied to NWP forecasts, provides a different than usual approach in the field of model verification, closer to the cluster-wise nature of the convective systems [24].

More complex algorithms may be introduced, already in use since last two decades for the optimal information extraction from multispectral and hyperspectral spaceborne data. The factor that changes substantially with hyperspectral data over that of conventional dimensionality is the magnitude of the volume available in this space. This is at once an opportunity and a challenge. The objective is to find a way to partition the N-dimensional feature space in such a way that the feature variations associated with the object of interest are contained within subregions of this space as uniquely and mutually exclusively as possible. The optimal methods are capable, in theory, of producing arbitrarily high accuracy even for classes that have a signal-to-background (or signal-to-noise) ratio that is very small, thus demonstrating the added value carried by such high-dimensional data [22].

A second track to foster next studies is the exploitation of Neural Networks (NN) for a couple of different, but complementary, topics. Firstly, the ability of NN-based algorithms for the signal extraction is a well documented field of research. Neural Networks can help the automatic classification of remotely sensed data. More important, the generalisation capabilities of this type of algorithms can fill the gaps in the performance of the current (deterministic) post-processing systems, nested downstream of NWP models and usually based on simple static-threshold classifiers [25].

Secondly, NN applications in the specific field of IR hyperspectral sounding have been proved as really beneficial, namely for the dimensionality reduction problem. As matter of fact, crucial decisions about the distribution of the huge volumes of data expected from IRS are currently undertaken by the EUMETSAT expert teams. In this perspective, the lossless or near-lossless

compression provided by the so-called autoassociative Neural Networks is recognized as a very effective technique [26].

## 7.4 Concluding remarks

---

The Infrared hyperspectral Sounder (IRS), planned on board the Meteosat Third Generation sounding mission in the 2022-23 timeframe, will be able to deliver unprecedented information on horizontally, vertically, and temporally (4-Dimensional) resolved water vapour and temperature structures of the atmosphere. It will provide extremely high spectral resolution and accurate measurements within the water vapour and carbon dioxide absorption bands, allowing the retrieval of finely resolved vertical structures of humidity (0.1–2 km resolution with 5–10% accuracy) and temperature (0.1–2 km with 0.5–1.5 K accuracy).

The analysis of the “big data” made available during this project by COMET and by EUMETSAT, trying to explore and exploit the potential applications of hyperspectral soundings for nowcasting of severe weather events was, of course, the main objective of this study. The results and the considerations reported here show that, even if further and deeper investigation is needed on this subject, it is possible – based on today’s knowledge – to draw a scientific (theoretical and experimental) baseline of the subject. In this perspective, the final outcome of this work is promising and future research activities can be easily addressed.

As additional and original contributions of the author to the research topic undertaken in this thesis, it is worth to mention, by one side, the involvement in the challenging and fruitful (at the same time) activities of the EUMETSAT expert groups, created to enhance the science and the user awareness about IRS. By the other side, the creation of the ItDP toolkit to handle and analyse the heterogeneous datasets needed to paint the scene with as more elements as possible represents – on the opinion of the author – a very useful by-product of the thesis work, envisaged to be exploited for future applications and to foster new generations of scientists to play with hyperspectral data [27].

# BIBLIOGRAPHY

---

- [1] EUMETSAT - MTG System Requirements Document (SRD). EUMETSAT Document EUM/MTG/SPE/06/0032, v4, 11 February 2011.
- [2] WMO - Statement of Guidance for Nowcasting, Very Short Range and Short Range Forecasting. Final Report of CBS Expert Team on the Evolution of Global Observing System, May 2012.
- [3] Paolo Bensi, Donny Aminou, J-L Bézy, Rolf Stuhlmann, Antonio Rodriguez, and Stephen Tjemkes. Overview of Meteosat Third Generation (MTG) activities. In *Second MSG RAO Workshop*, volume 582, page 37, 2004.
- [4] EUMETSAT - MTG Mission Requirements Document (MRD). EUMETSAT Document EUM/MTG/SPE/06/0011, v2C, 10 December 2007.
- [5] Antonio Rodriguez, Rolf Stuhlmann, Stephen Tjemkes, Donny M Aminou, Hendrik Stark, and Paul Blythe. Meteosat Third Generation: mission and system concepts. In *Infrared Spaceborne Remote Sensing and Instrumentation XVII*, volume 7453, page 74530C. International Society for Optics and Photonics, 2009.
- [6] EUMETSAT - MTG End-User Requirements Document (EURD). EUMETSAT Document EUM/MTG/SPE/07/0036, v3C, 30 March 2010.
- [7] Peter R. Griffiths and James A. De Haseth. *Fourier transform infrared spectrometry*, volume 171. John Wiley & Sons, 2007.
- [8] Domenico Solimini et al. *Understanding Earth Observation*. Springer, 2016.
- [9] EUMETSAT - Algorithm Theoretical Basis Document (ATBD) for Level 2 Processing of the MTG Infra-Red Sounder Data. EUMETSAT Document EUM/MTG/DOC/11/0188, v3, 26 February 2014.
- [10] Clive D. Rodgers. *Inverse methods for atmospheric sounding: theory and practice*, volume 2. World scientific, 2000.
- [11] J.R. Mecikalski, K.M. Bedka, and M. Koenig. Best practice document, 2012.
- [12] K. Lagouvardos, E. Liljas, and B. Conway. *Cost Action 78: improvement of nowcasting techniques : meteorology : final report*. COST action 78. Office for Official Publications of the European Communities, 2001.
- [13] CNMCA. *Manuale di meteorologia sinottica*. Servizio Meteorologico dell'Aeronautica Militare, 2006.
- [14] Bradley T Zavodsky, Shih-Hung Chou, Gary J Jedlovec, and William M Lapenta. The impact of near-real-time AIRS thermodynamic profiles on regional weather forecasting. In *Proc. Joint EUMETSAT Meteorol. Satell. Conf./15th Satell. Meteorol. Oceanogr. Conf. Amer. Meteorol. Soc.* EUMETSAT, 2007.
- [15] Hui Liu and Jun Li. An improvement in forecasting rapid intensification of Typhoon Sinlaku (2008) using clear-sky full spatial resolution advanced IR soundings. *Journal of Applied Meteorology and Climatology*, 49(4):821–827, 2010.

- 
- [16] Ralph A. Petersen, Robert Aune, Richard Dworak, and William Line. Using analyses of the information content of GOES/SEVIRI moisture products to improve very-short-range forecasts of the pre-convective environment. In *2012 EUMETSAT Meteorological Satellite Conference*, 2012.
- [17] Peter Bechtold. Atmospheric thermodynamics. *ECMWF Lecture Notes*, 2009.
- [18] Peter Bechtold. Atmospheric moist convection. *ECMWF Lecture Notes*, 2009.
- [19] Sangwon Joo, John Eyre, and Richard Marriott. The impact of MetOp and other satellite data within the Met Office global NWP system using an adjoint-based sensitivity method. *Monthly Weather Review*, 141(10):3331–3342, 2013.
- [20] P. Antonelli, A. Manzato, S. Puca, and F. Zauli. Evaluating atmospheric instability from high spectral resolution IR satellite observations. EUMETSAT Technical Report PA/IIS/FR/2010/01, 11 April 2011.
- [21] Jun Li, Chian-Yi Liu, Peng Zhang, and Timothy J. Schmit. Applications of full spatial resolution space-based advanced infrared soundings in the preconvective environment. *Weather and Forecasting*, 27(2):515–524, 2012.
- [22] David Landgrebe et al. Multispectral data analysis: A signal theory perspective. *Purdue Univ., West Lafayette, IN*, 1998.
- [23] Stephen Tjemkes, E. Gregow, A. Dybbroe, T. Leppelt, A. Vocino, M. Martinez, S. De Haan, M. Irsic, and R. Stuhlmann. MTG-IRS Near Real Time Demonstration Project. In *Proceedings of the 2016 EUMETSAT Meteorological Satellite Conference, Darmstadt, Germany*, 2016.
- [24] Christopher A. Davis, Barbara G. Brown, Randy Bullock, and John Halley-Gotway. The method for object-based diagnostic evaluation (MODE) applied to numerical forecasts from the 2005 NSSL/SPC Spring Program. *Weather and Forecasting*, 24(5):1252–1267, 2009.
- [25] Fabio Del Frate, Fabio Pacifici, Giovanni Schiavon, and Chiara Solimini. Use of neural networks for automatic classification from high-resolution images. *IEEE transactions on geoscience and remote sensing*, 45(4):800–809, 2007.
- [26] Giorgio A. Licciardi and Fabio Del Frate. Pixel unmixing in hyperspectral data by means of neural networks. *IEEE transactions on Geoscience and remote sensing*, 49(11):4163–4172, 2011.
- [27] Antonio Vocino, Fabio Del Frate, and Giovanni Schiavon. Use of MTG-IRS proxy data for early detection of convective systems. In *Proceedings of the 2017 EUMETSAT Meteorological Satellite Conference, Rome, Italy*, 2017.

# LIST OF FIGURES

---

2.1	Artist's rendition of the MTG-S (in the foreground) and MTG-I series spacecraft in orbit (image credit: ESA). . . . .	11
2.2	MTG missions and main user applications beneficiaries (from [5]).	12
2.3	MTG missions spatial coverage and temporal rate (from [5]). .	13
2.4	MTG in orbit deployment scenario. A temporal shift of the whole programme is planned, thanks to the extension of the operational life of MSG satellites <sup>22</sup> (image credit: EUMETSAT).	13
2.5	MTG system segments and operational sites (from [5]). The overall system is divided in five segments: Space, Launch and Early Operation, Core Ground Segment, Application Ground Segment, User and Operations Support (left). The system components will be distributed over various sites to pursue the highest resilience. Commercial communication satellite services will be used, as currently in the EUMETCAST system, for dissemination of data to regional and global users (right). . . .	14
2.6	Illustration of the IRS instrument (preliminary design of Kayser-Threde GmbH). . . . .	16
2.7	Spectral coverage of MTG-IRS, identifying the two bands of the instrument, compared to METOP-IASI coverage spanning the whole wavenumber range ( $645 - 2760 \text{ cm}^{-1}$ ) shown in the figure (image credit: EUMETSAT). . . . .	16
2.8	Local Area Coverage scanning zones, as from [6] (image credit: EUMETSAT). . . . .	17
2.9	Details of LAC zone 4 (image credit: EUMETSAT). According to [6], the polygonal line AH defines the mandatory lower boundary of LAC4. . . . .	17
2.10	Operational practice for the IRS scanning sequence, as from [6] (image credit: EUMETSAT). . . . .	18
2.11	High level user requirements for the level 2 products derived from MTG-IRS observations (from [6]). . . . .	19
3.1	Example of use of satellite-derived instability index for nowcasting purposes at UKMO (image courtesy of Ed Pavelin). . . . .	25
3.2	Example of comparison between MSG-SEVIRI to MTG-IRS for the detection of instability at CIMSS (image courtesy of Ralph Petersen). . . . .	26
3.3	(a) Observation impact and (b) mean impact per observation of the instruments using MWS and IRS techniques (from [19]). . .	27

3.4	Prediction of instability over Italy using IASI data (image courtesy of Paolo Antonelli and Agostino Manzato). Training-Validation (TV) diagram of the Cross-Entropy Errors (CEEs) computed over multiple instances of training/validation sets to infer the prediction skill of the Neural Networks needed to select the best predictors of the classification algorithm (top-left). Region of interest for the study (top-right). Contingency table scores for Instability Indices derived from radiosoundings and IASI retrievals compared with lightning occurrences (bottom-left). Time series of the CAPE indices computed from radiosoundings and from IASI retrievals, overlapped with lightning coincident events (bottom-right). . . . .	28
3.5	Baseline Data Processing – NWP component. . . . .	31
3.6	Baseline Data Processing – observation component. IASI (top-left) and radiosounding (bottom-right) assimilated profiles, with associated impact on objective analysis. Italian weather radar (top-right) and lightning (bottom-left) networks. . . . .	32
4.1	Schematic picture of ICT assets for data processing and distribution of the COMET Centre. . . . .	37
4.2	Job scheduler of ItDP processor for MTG-IRS proxy data, showing both online and offline batch queues. . . . .	38
4.3	Example of IASI historical data downloaded from EUMETSAT EO portal and processed by ItDP processor. . . . .	39
4.4	Example of map of IASI instability index (CAPE) computed by ItDP processor for a case study. . . . .	40
4.5	Example1 of Skew-T diagram of IASI data computed by ItDP processor. . . . .	40
4.6	Example2 of Skew-T diagram of IASI data computed by ItDP processor. . . . .	41
4.7	Example of distribution of IASI instability index (CAPE) computed by ItDP processor. . . . .	41
4.8	Example of map of IRS-proxy instability index (CAPE) computed by ItDP online processor. . . . .	42
4.9	Example of Skew-T diagram of IRS-proxy data computed by ItDP online processor. . . . .	43
4.10	Another example (CrIS as input) of map of IRS-proxy instability index (CAPE) computed by ItDP offline processor. . . . .	43
5.1	Snapshot of the git interface used in this project for track changes and software optimisation. . . . .	47
5.2	Functional blocks of the main procedures involved in the ItDP software system. . . . .	48



5.3	Procedures – Processing System. . . . .	49
5.4	Procedures – Supervisor System. . . . .	49
5.5	Procedures – External Libraries. . . . .	50
5.6	Modules of GRID procedure. . . . .	50
5.7	Extract from IRS Level 2 Atmospheric State Product element list (from [9]). . . . .	53
5.8	Extract from IRS Level 2 Atmospheric State Product element list (continued). . . . .	54
5.9	Extract of source code from the Object Detection procedure. . . . .	55
5.10	Extract from the code profiler output for IRS end-to-end processor implemented in ItDP. Connections indicate the relationships across the software modules. Hierarchical structure (top-bottom) and execution times (percentage of total time and number of calls) of the individual modules are shown. . . . .	59
5.11	Extract from the code profiler output for IRS end-to-end processor implemented in ItDP (detail). . . . .	59
5.12	Example of end-to-end processor, showing the main functional blocks identified as ItDP procedures. Supervisor (blue) and processing (yellow) blocks, as well as external libraries (green) have been interconnected to realise a stand-alone, cascading processor for IRS proxy and auxiliary data. . . . .	60
6.1	Extract from list of significant convection case studies. . . . .	64
6.2	Sequence of composite images (IR satellite-radar-lightning data) for 16 Apr. 2017 case study. Top-left: obs+0h, top-right: obs+3h, bottom-left: obs+6h, bottom-right: obs+9.3h. . . . .	66
6.3	Map of IRS-proxy CAPE instability index for 16 Apr. 2017 case study. . . . .	67
6.4	Map of IRS-proxy cloud mask for 16 Apr. 2017 case study. . . . .	67
6.5	Distribution of IRS-proxy CAPE instability index for 16 Apr. 2017 case study. . . . .	68
6.6	Spatial-temporal colocation of IRS-proxy data and observed event for 16 Apr. 2017 case study. Left: event at 19.00 UTC, right: observation at 09.40 UTC. . . . .	69
6.7	Sequence of 6 hours-step probability forecasts of COSMO model for 16 Apr. 2017 case study. . . . .	69
6.8	Significant weather map (post-processing product) from deterministic forecast of COSMO model for 16 Apr. 2017 case study. . . . .	70
6.9	Example 1 of revisiting time for 7-8 May 2017 case study – proxy data. . . . .	71
6.10	Example 1 of revisiting time for 7-8 May 2017 case study – event. . . . .	72

6.11	Example 2 of revisiting time for 7-8 May 2017 case study – proxy data. . . . .	72
6.12	Example 2 of revisiting time for 7-8 May 2017 case study – event. . . . .	73
6.13	Composite image (IR satellite-radar-lightning data) for 6 May 2017 case study. . . . .	74
6.14	Map of IRS-proxy CAPE instability index for 6 May 2017 case study. . . . .	74
6.15	Zoom of the map of IRS-proxy CAPE instability index for 6 May 2017 case study. . . . .	75
6.16	Example of object detection by spatial clustering for 6 May 2017 case study. . . . .	76
6.17	Map of IRS-proxy CAPE instability index for 10 April 2017 case study - NRT data. . . . .	77
6.18	Map of IRS-proxy CAPE instability index for 10 April 2017 case study - NRT a priori data. . . . .	78
6.19	Map of IRS-proxy CAPE instability index for 10 April 2017 case study - reprocessed data. . . . .	79
6.20	Map of IRS-proxy CAPE instability index for 10 April 2017 case study - reprocessed a priori data. . . . .	80
7.1	Comparison of posterior (left) and prior (right) distribution of CAPE for the same observed IRS proxy granule. . . . .	82
7.2	Detail of the colocation of IRS-proxy data and observed event for 16 Apr. 2017 case study. Left: event at 19.00 UTC, right: observation at 09.40 UTC. . . . .	83
7.3	Detail of revisiting time feature for 7-8 May 2017 case study – proxy data and event. . . . .	84

# LIST OF TABLES

---

2.1	MTG overall timeline . . . . .	10
5.1	List of formats handled by the Data Input procedure. . . . .	52
5.2	List of main 3rd party libraries used by ItDP toolkit. . . . .	58

