H2020-EINFRA-2015-1

VI-SEEM

VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean

Deliverable D6.2

1st Report of open calls and integration support

Author(s): A. Athenodorou (editor)

Status –Version: Final – j

Date: July 31st, 2018

Distribution - Type: Public

Abstract: Deliverable D6.2 – "1st Report of open calls and integration support" – this is report on the implementation of the 1st call, the outcome of the review process and the required efforts for the integration of the selected applications in the VRE. After the informal check, an update was requested to include results of the second call. Final data will be provided in D6.4 as per DoA.

The VI-SEEM Consortium consists of:

GRNET	Coordinating Contractor	Greece
CYI	Contractor	Cyprus
IICT-BAS	Contractor	Bulgaria
IPB	Contractor	Serbia
NIIF	Contractor	Hungary
UVT	Contractor	Romania
UPT	Contractor	Albania
UNI BL	Contractor	Bosnia-Herzegovina
UKIM	Contractor	FYR of Macedonia
UOM	Contractor	Montenegro
RENAM	Contractor	Moldova (Republic of)
IIAP-NAS-RA	Contractor	Armenia

GRENA	Contractor	Georgia
BA	Contractor	Egypt
IUCC	Contractor	Israel
SESAME	Contractor	Jordan

The VI-SEEM project is funded by the European Commission under the Horizon 2020 e-Infrastructures grant agreement no. 675121.

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
June 13th, 2017	a	Ioannis Liabotis, Andreas Athenodorou	TOC and initial contents provided
June 23rd, 2017	b	Andreas Athenodorou	Expanded various sections
June 26th, 2017	c	Ioannis Liabotis	Various corrections and editing
June 29th, 2017	d	Andreas Athenodorou, Lidija Milosavljevic	Applied the quality control changes and various corrections and editing.
July 6th , 2018	e	Andreas Athenodorou	Include information for the 2nd Open Call
July 20th, 2018	f	Evangelia Athanasaki	Include information on the results for the 1st Open Call
July 24th, 2018	h	Andreas Athenodorou, Evangelia Athanasaki, Dusan Vudragovic	Editing
July 30th, 2018	i	Andreas Athenodorou, Lidija Milosavljevic	Applied the quality control changes and various corrections and editing.
July 31st, 2018	j	Ognjen Prnjat	Final corrections and additions
July 31st, 2018	k	Andreas Athenodorou	Final corrections and additions

Table of contents

References

[1] Project VI-SEEM-675121 - Annex I - Description of the Action

List of Figures

List of Tables

Glossary

API	Application Programming Interface
CBIR	Content-Based Image Retrieval
CORDEX	Coordinated Regional Downscaling Experiment
CRYSTAL	Computational tool for solid state chemistry and physics
DDS	Drug Delivery System
DICOM	Digital Imaging and Communications in Medicine
DREAM	The Dust REgional Atmosphere Model
ECHAM	Global Climate Model developed by the Max Planck Institute for Meteorology
ECMWF-ERA	European Centre for Medium-Range Weather Forecast – Ranalysis
EMAC	ECHAM/MESSy Atmospheric Chemistry
Ferret	Interactive Computer Visualization and Analysis environment
GrADS	Grid Analysis and Display System
GROMACS	Molecular Dynamics Software Toolkit
HPC	High Performance Computing
HR	Health Risks
ICT	Information Communication Technology
NAMD	Scalable Molecular Dynamics Toolkit
NWP	Numerical Weather Prediction
OCR	Optical Character Recognition
QL	Quality of Life
RegCM	The Regional Climate Model system
SDDN	Smart Drug Delivery Nanocarriers
SDS-WAS	Sand and Dust Storm Warning Advisory and Assessment System
SEEM	South East Europe and Eastern Mediterranean region
SLES	Sodium Lauryl Ether Sulphate
VRE	Virtual Research Environment
VI-SEEM	VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean
WCRP	World Climate Research Project
WRF	The Weather Research & Forecasting Model
WMO	World Meteorological Organization

Executive summary

What is the focus of this Deliverable?

The focus of this deliverable is to provide a report on the implementation of the 1st and 2nd calls for proposals for projects accessing VI-SEEM resources and services, the outcome of the review process and the required efforts for the support of the selected applications. In addition we provide the production of results for the 1st open call within the VI-SEEM Virtual Research Environment (VRE). Results obtained through the 2nd open call will be provided in the deliverable D6.4.

Furthermore we provide a detailed description of the applications coming from the three scientific communities of Climate, Digital Cultural Heritage and Life Sciences. Moreover, a detailed flow chart which demonstrates how the two calls have been implemented is provided. This deliverable sets the ground for the implementation of the next open call for applications.

After the informal check, an update of D6.2 was requested by the reviewers to include results of the second call. Final data will be provided in D6.4 as per DoA.

What is next in the process to deliver the VI-SEEM results?

The documentation presented in this deliverable will be used as a basis for providing the final deliverable D6.4 as per DoA.

What are the deliverable contents?

The deliverable "1st Report of open calls and integration support" provides a description of the implementation of the 1st and 2nd calls for proposals of projects accessing VI-SEEM resources and services in detail. It consists of all the information relevant to their implementation. The description starts with a short introduction on the basics details of the calls and more specifically of the 1st and 2nd calls. Subsequently, Section 2 provides a short summary on the adopted procedure and the objectives of the calls. Then, Section 3 provides all the information on the available computational and storage resources as well as services. Section 4 gives the list of all the applications that have been received in the two calls as well as description of the review process. Section 5 provides the outcome of the review procedures as well as the plans for supporting the deployment and the operation of the applications. Section 6 gives a description of the scope of the selected applications. Section 7 provides a description of the results obtained in the 1st Open Call as well as information on the publications, presentations as well as PhD and MCs thesis resulted within this call. Finally, Section 8 provides the conclusion of the deliverable.

Conclusions and recommendations

This deliverable provides details regarding the implementation of the 1st and 2nd calls for proposals for projects accessing VI-SEEM resources and services. The goal of these calls is the capacity building by producing fresh results which will be published as well as by enriching further the VRE portal with datasets and codes.

The 1st and 2nd calls have been organized according to the outcome of deliverable D6.1 “Framework for VRE resource and service provision” and was addressed to researchers working in the scientific fields of Life Sciences, Climate Research, and Digital Cultural Heritage.

The first call enabled 20 scientific applications to get access to the VI-SEEM resources and services. From the applications 11 belong to Climate Research, 5 to Life Sciences and 4 to Digital Cultural Heritage. The second call enabled 17 scientific applications to get access to the VI-SEEM resources and services. From the applications 4 belong to Climate Research, 7 to Life Sciences and 6 to Digital Cultural Heritage.

VI-SEEM consortium members assigned service enablers for each participant - the role of which guide researchers to access the allocated resources and services and advise the progression of the projects. The access to the underlying computational resources was awarded for a maximum period of 12 months and access to the underlying storage resources can be provided for up to 2 years, beginning January 2017 and September 2017.

1 Introduction

This deliverable describes the setup and implementation of the 1st and 2nd calls for proposals for projects accessing VI-SEEM services. These two calls belong to the series of 3 open calls for projects that require the use of VI-SEEM services and resources in order to create scientific output relevant to the three communities in the South Eastern and Eastern Mediterranean region, namely Climate Research (CR) community, Life Sciences (LS) community and Digital Cultural Heritage (DCH) community.

The design and implementation of the 1st and 2nd calls has been based on the outcome of deliverable D6.1 “Framework for VRE resource and service provision” where the main principles and objectives of the VI-SEEM service and resources access framework have been developed.

The defined framework is being used in the context of the VI-SEEM project for providing access to users of the VI-SEEM services and the associated resources. It provides the foundation for the access policies that are used throughout the duration of the project to ensure the best possible utilization of the VI-SEEM services and resources, serving also as a long-term guide for the provision of Virtual Research Environment (VRE) services.

The VI-SEEM VRE is composed of several types of services ranging from access to multiple types of computational services such as HPC, Cloud and Grid, storage services and underlying resources, as well as software, data, workflow repositories and application specific services. All those services have different characteristics in terms of availability, costs and therefore access modalities.

The service provision framework for the VRE resource and service provision is based on the following principles:

- To make services and resources provided by the VRE available for free to as many as possible researchers from the region of South Eastern Europe and the Eastern Mediterranean (SEEM) and the wider user community (for open and free resources).
- To promote the best and most efficient usage of the underlying infrastructure provided by the VI-SEEM resource provider partners.
- To promote scientific excellence and improve the competitiveness of the researchers in the SEEM region.
- To promote scientific collaboration and exchange of knowhow between the experienced research groups and the less experienced but potentially excellent new research teams of the region.
- To open up the knowledge and data produced in the region to all researchers in Europe and beyond where possible.
- To offer user and application enabling support to user communities which require it.
- To serve the scientific fields of Life Sciences, Climate Research and Digital Cultural heritage which are identified as most relevant for the region.

- To provide the opportunity to researchers of all countries to have access to the services offered by VI-SEEM.

All the above principles have been used to implement the 1st and 2nd VI-SEEM calls.

Section 2 provides a summary of the procedure and the objectives of the calls. Section 3 provides all the information on the available computational and storage resources as well as services. Section 4 gives the list of all the applications that have been received in the two calls as well as description of the review process. Section 5 provides the outcome of the review procedures as well as the plans for supporting the deployment and the operation of the applications. Section 6 gives a description of the scope of the selected applications. Section 7 provides a description of the results obtained in the 1st Open Call as well as information on the publications, presentations as well as PhD and MCs thesis resulted within this call. Finally, Section 8 provides the conclusion of the deliverable.

2 Summary of the process and objectives of the calls

VI-SEEM offers a broad set of generic as well as application-specific services in the region of South-eastern Europe and the Eastern Mediterranean, with special focus on the scientific communities of Life Sciences, Climatology and Digital Cultural Heritage. Such services are in the areas of Compute resource provisioning (HPC, Grid and Cloud), Storage and Data services provisioning, Dataset provisioning, Software and Scientific Workflow provisioning as well as Application Specific service provisioning. These services create a unique Virtual Research Environment (VRE), thus improving research productivity and competitiveness on the pan-European level.

The calls were addressed to scientists and researchers that work in academic and research institutions in the region of South Eastern Europe and the Eastern Mediterranean. More specifically these are (in alphabetical order): Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Cyprus, Egypt, FYR of Macedonia, Georgia, Greece, Hungary, Israel, Jordan, Lebanon, Moldova, Montenegro, Romania, Serbia and Turkey.

The project proposals should address open research topics in specific fields of Life Sciences, Climate research, and Digital Cultural Heritage.

Via these calls VI-SEEM opened possibilities for regional scientists from the selected scientific fields to have access to the advanced resources and services that it offers.

The list of services and resources offered by the VI-SEEM VRE can be found at: https://services.vi-seem.eu and in the VI-SEEM VRE at: https://vre.vi-seem.eu.

Access to underlying computational resources can be awarded for a maximum period of 12 months, while access to underlying storage resources may be provided for up to 2 years.

The first and second calls enables researchers from selected countries and research fields to obtain access to the advanced services of the VI-SEEM Virtual Research Environment.

2.1 Applicable scientific fields

Eligible projects are only the ones that address one of the following scientific and/or societal challenges:

In the field of Life Sciences

- LS Area A: Modelling and Molecular Dynamics (MD) study of important drug targets.
- LS Area B: Computer-aided drug design.
- LS Area C: Analysis of Next Generation DNA sequencing data.
- LS Area D: Synchrotron data analysis.

- LS Area E: Image processing for biological applications.

In the field of Climate Research

- CR Area A: Regional climate modelling to better understand and predict climate change and impacts, and phenomena such as dust storms.
- CR Area B: Air quality modelling, including atmospheric chemistry and air pollution transport.
- CR Area C: Weather forecast and extreme weather prediction, model development, application.

In the field of Digital Cultural Heritage

- DCH Area A: Online services and access to repositories in order to enable studies of the cultural heritage assets in the region (e.g., searchable digital libraries; with support of meta-data and OCR for Latin characters).
- DCH Area B: Online visualization tools and data management systems to drive breakthrough contributions to art historical problems (e.g., interactive visualization viewer of RTI files and 3D models with digital libraries integration).
- DCH Area C: Unsupervised feature learning in photogrammetric techniques, data processing for image classification; semantic referencing; and geo-referencing.

The criteria for the evaluation of projects for accessing the available resources where:

- Scientific excellence.
- Scientific and/or social impact of the proposed research.
- The need for usage of the selected services and resources.
- The ability to provide project results (mainly data sets but also services and software) as services for other future VRE users.
- Maturity and experience of the principal investigator and his/her team in the research field as well as in using the selected resources and services.
- Feasibility of the project based on the technical evaluation and the availability of resources.
- Potential for the collaboration among scientists in more than one eligible country for this call.

The criteria above where used for the evaluation of proposals at the review process.

VI-SEEM aims at a balanced provision of resources to the whole spectrum of scientific fields between the three target communities that this call addresses, as well as to as many as possible countries in the South-eastern Europe and Easter Mediterranean region.

2.2 Eligibility

Eligible applicants (as Principal Investigators) are scientists affiliated with academic or research institutions in the following countries (in alphabetical order): Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Cyprus, Egypt, Former Yugoslav Republic of Macedonia, Georgia, Greece, Hungary, Israel, Jordan, Lebanon, Moldova, Montenegro, Romania, Serbia and Turkey.

Collaborators in proposals might reside in any country provided that no specific geographical restrictions apply for access by the corresponding centres that offer resources in the various resource-providing countries.

Industrial partners may participate only as collaborators in proposal that is led by academic or research institutions in the eligible countries, and only if the aims and objectives of the project is open research with results to be published in research journals or conferences.

Applicants should commit to using the resources that will be allocated to them, as well as to providing reports of their work based on the proposed time schedule (see below). Further to that, scientists should acknowledge the use of the VI-SEEM VRE services in all publications presenting results obtained from using the allocated resources.

2.3 Application process

All proposals have been submitted electronically via the VI-SEEM survey tool.

The application form was also available in a pdf format in order for applicants to have the full list of questions available. The applicants had to fill in the on line form for their applications to be taken into account.

Support to applicants was provided during the two calls: namely, VI-SEEM Access Team was available to answer questions while a call was open.

2.4 Important dates

2.4.1 Important dates for the 1st Open Call

In the following list we provide all the important dates concerning the 1st call of applications for access to VI-SEEM e-Infrastructure resources and services.

- Opening date: 14th of October 2016.

- Closing Date: 14th of November 2016.
- Clarification provided by applicants if needed: 18th of November 2016.
- Allocation decision: January 2017.
- Allocation Start Date of awarded proposals: January 2017.
- Allocation end date of award: January 2018 for computational services and January 2019 for some data services.
- Final report from successful projects: May 2018.

2.4.2 Important dates for the 2ndt Open Call

In the following list we provide all the important dates concerning the 2nd call of applications for access to VI-SEEM e-Infrastructure resources and services. (Dusan Please Update the following dates)

- Opening date: 4th May 2017.
- Closing Date: 26th June 2017.
- Clarification provided by applicants if needed: 31st July 2017.
- Allocation decision: 1st September 2017.
- Allocation Start Date of awarded proposals: 30th September 2017.
- Allocation end date of award: September 2019 for computational projects, September 2020 for some data projects.
- Final report from successful projects: November 2019.

3 Available resources and services

The infrastructure of the VI-SEEM VRE consists of resources of various types – HPC resources – clusters and supercomputers with different hardware architectures, Grid sites, Clouds with possibility to launch virtual machines (VMs) for services and distributed computing, and storage resources with possibility for short and long term storage.

3.1 Computational and storage services available

HPC resources: The HPC resources of the project consist of clusters with low-latency interconnection or supercomputers. Most of the systems are based on CPUs with x86_64 instruction set, some of them equipped with accelerators (GPUs and Intel Xeon Phi coprocessors), but also there are BlueGene/P systems, as well as one based on the Cell processor (PS3 cluster IMAN1-Booster/King). HPC resources are offered by the following countries: Albania, Armenia, Bulgaria, Cyprus, Egypt, FYR of Macedonia, Greece, Hungary, Jordan, Romania and Serbia. In total 19 million CPU core hours, 370 million GPU core hours and 16 million Phi core hours will be provided in this call. Information regarding the HPC resources can be found in the following link:
https://wiki.vi-seem.eu/index.php/Main_Page#HPC_Resources

Cloud resources: The Cloud resources available in the call can be used in two ways. Those clouds that provide the ability to launch VMs with public IPs give the possibility to deploy VRE services for their main or backup/fail-over instance. VMs that possess only private IPs can be used for distributed data processing where necessary. Cloud resources are provided by the following countries: Albania, Armenia, Bosnia and Herzegovina, Bulgaria, Cyprus, FYR of Macedonia, Greece, Israel, Moldova and Romania. In total around 300 VM cores are to be provided in this call. Information on the Cloud resources can be found in the following link:
https://wiki.vi-seem.eu/index.php/Main_Page#Cloud_Resources

Grid resources: The Grid resources, available in this call, are provided mostly from smaller clusters. Grid resources for the VI-SEEM VRE are provided from the following countries: Armenia, Bulgaria, FYR of Macedonia, Georgia, Greece, Moldova, Montenegro and Serbia. Information on the Grid resources can be found in the link:
https://wiki.vi-seem.eu/index.php/Main_Page#Grid_Resources

VI-SEEM Simple Storage service: The VI-SEEM Simple Storage service (VSS) is a secure data storage service provided to VI-SEEM users for storing and sharing research data as well as keeping it synchronized across different computers. Data sharing will be possible with other registered VI-SEEM users or with anyone else by using public links which can be protected with passwords if needed. Each user will be provided with 50 GB of storage for up to two years from the beginning of their project.

VI-SEEM Repository Service: The main storage service that will allow the users of the VI-SEEM VRE to deposit and share data is the VI-SEEM Repository Service (VRS). This is VI-SEEM main repository for hosting the Regional Community Datasets. It can also be used to host publications and their associated data, as well as software or references to software and workflows used to generate such data and publications. The VRS is also the service designated for storing simplified data formats such as images, videos or others formats suitable also for the general public. Each project is eligible for up to 10 TB of storage for up to two years from the start of the project.

VI-SEEM archival service: Data archiving is the practice of moving data that is no longer being used or are being used in a less frequent fashion into a separate storage service. It is a single set or a collection of historical records specifically selected for longer term retention and future reference. Additionally, data archives contain data that are important for future reference or it is important to preserve them for regulatory and audit purposes. In science, archived data are important for future reference and reproducibility of scientific simulations. Each project will be eligible for storing up to 10 TB in the archival service of the project for up to two years from the start.

VI-SEEM work storage space / local storage and data staging: This service refers to storage space available by the computational resource providers to store temporary data for the purposes of processing them, or for storing results of computations. The service will be available for 12 months from the start of the project. The maximum capacity depends on the service provider.

3.2 Application-specific services

3.2.1 Application-specific services for Climate

Live Access Server (LAS)

The Live Access Server is a highly configurable server designed to provide flexible access to geo-referenced scientific data. It can present distributed data sets as a unified virtual data base through the use of DODS networking. Ferret is the default visualization application used by LAS, though other applications (Matlab, IDL, GrADS etc) can also be used.

LAS enables the web user to:

- Visualize data with on-the-fly graphics.
- Request custom subsets of variables in a choice of file formats.
- Access background reference material about the data (metadata).
- Compare (difference) variables from distributed locations.

3.2.2 Application-specific services for Digital Cultural Heritage

VI-SEEM Clowder

Clowder is a research data management system designed to support any data format and multiple research domains. It contains three major extension points: pre-processing, processing and previewing. When new data is added to the system, pre-processing is off-loaded to extraction services for extracting appropriate data and metadata. The extraction services attempt to extract information and run pre-processing steps based on the type of the data, for example to create previews. This raw metadata is presented to the user in the Clowder web interface. Users can upload, download, search, visualize and get various information about these data.

Data in the case of VI-SEEM and more specifically in the field of Digital Cultural Heritage can be of very diverse types and formats.

More specifically users can upload massively (zipped) or individual files of:

- 3D Models: where extractors clean up and prepare for visualization on the platform itself.
- Scanned books and their metadata: OCR algorithms will be used to extract the text in the documents so that users can find books using both metadata information and the book’s contents.
- Image, text and sound files and their metadata, organised in collections.
- Advanced documentation data, such as Reflectance Transformation Imaging, and analysis of material properties of structures, works of art and artefacts.

VI-SEEM CHERE

CHERE (Cultural HEritage REpository) is a collection of individual applications and services aimed at simplifying the handling of digital cultural heritage resources. It currently consists of module for repository and management of entries (based on Omeka-S), module for structure-from-motion reconstruction (based on VisualSFM, vlfeat,PMVS/CMVS, SPSR and texrecon) provided as both hosted web service as well as semi-automated Docker container, measurement module and web based module for basic editing of 3D meshes (Meshlab.js). It also provides basic tools to generate web based panoramas from 360 photos (pannellum).

This service significantly lowers the required technical knowledge required to start using digital repositories and conduct structure from motion reconstructions without installation of additional software. It is aimed at museum staff with limited technical know-how but can also be used in standalone version via Docker containers for more technically inclined users. It can generate tiled models of panoramas suitable for web consumption without the restriction of maximum resolution supported by the browser.

VI-SEEM ARCHES (MC4CH)

MC4CH (Management Cloud for Cultural Heritage) is a service which significantly lowers the required technical knowledge required to start using digital repositories and conduct structure from motion reconstructions without installation of additional software. It is aimed at museum staff with limited technical know-how but can also be used in standalone version via Docker

containers for more technically inclined users. It can generate tiled models of panoramas suitable for web consumption without the restriction of maximum resolution supported by the browser.

3.2.3 Application-specific services for Life Science

ChemBioServer

ChemBioServer is a web-application for effectively mining and filtering chemical compounds used in drug discovery. ChemBioServer allows for pre-processing of compounds prior to an in silico screen, as well as for post-processing of top-ranked molecules resulting from a docking exercise with the aim to increase the efficiency and the quality of compound selection that will pass to the experimental test phase.

It provides researchers with the ability to:

- Browse and visualize compounds along with their properties.
- Filter chemical compounds for a variety of properties such as steric clashes and toxicity.
- Apply perfect match substructure search.
- Cluster compounds according to their physicochemical properties providing representative compounds for each cluster.
- Build custom compound mining pipelines.
- Quantify through property graphs the top-ranking compounds in drug discovery procedures.

AFMM

AFMM provides an automated platform with which the users can generate parameters for modelling small molecules with Molecular Dynamics simulations. The method used fits the molecular mechanics potential function to both vibrational frequencies and eigenvector projections derived from quantum chemical calculations. The program optimizes an initial parameter set (either pre-existing or using chemically-reasonable estimation) by iteratively changing them until the optimal fit with the reference set is obtained. By implementing a Monte Carlo-like algorithm to vary the parameters, the tedious task of manual parameterization is replaced by an efficient automated procedure. The program is best suited for optimization of small rigid molecules in a well-defined energy minimum, for which the harmonic approximation to the energy surface is appropriate for describing the intra-molecular degrees of freedom.

Due to the abundance of organic molecules, no parameters have been created for the full chemical space. Thus, there is a great need for molecule parameterization before proceeding to Molecular Dynamics calculations. AFMM allows users to access parameters for their Molecular Dynamics simulation of small organic molecules that can be used as drugs or materials.

NANO-Crystal

NANO-Crystal is a web-based tool, is implemented for the construction of spherical nanoparticles of a given radius.

More specifically, the goal is to find the number and the Cartesian coordinates of smaller spheres that fit on the surface of the nanoparticle and visualize the output morphology. The home page (http://nanocrystal.vi-seem.eu/) menu allows two selections for the user:

i. the radius of the nanosphere (nm), and
ii. the radius of smaller spheres (nm), that will cover the surface of the nanoparticle

The program computes the number of smaller spheres that fit on the bigger surface and the user can download their Cartesian coordinates (output format .xyz). The program code is implemented using PHP server-side scripting language, which is embedded into the HTML and CSS code. JQuery, a cross-platform JavaScript library, is also used. For local host of the webpage tool, the Wamp server is used. Moreover, we have developed a crystal computational morphology toolbox for constructing and modelling different crystal nanoparticle shapes. We use computational approaches for computing the macroscopic morphology of any periodic crystal by forming different shapes based on Miller indices and the distance measure from the centre of the crystal and visualizing the resulting crystal. That crystal is a polyhedron that is created as the intersection of multiple polyhedra and individual planes via the steps that follows. This tool is planned to be imported in the NANO-Crystal webserver within 2017.

This tool enables users to construct spherical nanoparticles as well as different crystal nanoparticle shapes based on Miller indices and the distance measure from the centre of the crystal.

Subtract

Subtract is an online tool that can calculate the volume of a binding site found in a protein. Subtract accepts an atom selection in the form of a PDB file and computes the three-dimensional convex hull of the atoms points with the help of SciPy library. The next step of the algorithm is to compute the volume of the convex hull and the volume of the atoms that are included in the solid based on their van der Waals radii. The subtraction of those two volumes yields the volume of the investigated cavity. The algorithm computes cavity volumes of trajectory frames in parallel for maximum efficiency and speed. It requires minimal usage of memory due to the fact that it follows a buffering strategy of reading file chunks and therefore there is no need to load the entire file into memory. There is a wide support of trajectory formats like Gromacs trajectory files and multi-model PDB files due to its dependency to the MDTraj library.

The measurements are evaluated for statistical significance using Wilcoxon Signed-Rank test and had their null hypothesis rejected (p-value < 0.005). Subtract is a tool that has been created to solve the problem of accurate measurement of the protein binding sites, and works

both for crystal structures downloaded from the Protein Data Bank and for protein structures arising from Molecular Dynamics simulations trajectories.

DICOM

"DICOM Network" provides access to investigations for medical staff with the appropriate access rights and as well as patients to the personal radiography investigations. Nowadays the system collects and processes more than 500 gigabytes of data per month. The system is based on Data Storage and Data Processing components distributed between different processing units and storages, which could be customized using specific interfaces. VI-SEEM integration connects national DICOM Network application, that it is containing existing DICOM Portal http://dicom.md/, with the DICOM Portal installed on VI-SEEM platform resources. DICOM DATA Interface grants the interconnectivity for different users of the both portals and allows displaying DICOM investigations using both portals interfaces. Public DICOM Server grants possibility for any VI-SEEM platform member to pull and retrieve the investigations from DICOM Network application and use the developed facilities based on configured access rules. VI-SEEM platform will offer possibility to install and configure publicly available DICOM Portal that can be used by any interested institutions to store, access and share medical images. Setting up public DICOM Portal instance will increase the level of access to DICOM investigations and will help to promote DICOM Network services to regional medical research and practicing community.

3.2.4 Assignment of Resources

Finally, successful applicants will get access to all public services, data sets, workflows and codes available in the VI-SEEM VRE – https://vre.vi-seem.eu.

The number of accepted projects depends on the technical and scientific merit of the proposals and the availability of resources. HPC projects were expected to be assigned a maximum of 3 Million core hours. A larger number of requested core hours requires elaborate and well documented justification.

4 Applications and review

4.1 Applications from the 1st Open Call

In total 22 applications have been received in this call from 10 different countries of the region. 14 of the applications required HPC services, 6 Grid and Cloud service, 12 of storage services, 8 application specific services. The applications underwent technical review and where all the requirements were clarified they were sent for the scientific review. For confidentiality reasons the following tables show only the accepted applications.

Application Acronym	Scientific Discipline	Country	HPC	Grid	Cloud	Simple Storage	Reposi tory	Archive
VINE	CR Area A	Georgia	Y	Y			Y	
PSOMI	LS Area A	Montene gro			Y			
MECCA_ECHAM_G PU	CR Area A	Cyprus	Y					
TVRegCM	CR Area A	Bulgaria	Y			Y	Y	
ACIQLife	CR Area B	Bulgaria	Y			Y	Y	
CH-CBIR	DCH Area C	Bosnia	Y					
Continuous_LST	CR Area A	Israel			Y	Y	Y	
MULTIDRUG	LS Area A	Bulgaria	Y					
DATANMNH	DCH Area A	Bulgaria			Y			
D3R	LS Area B	Greece	Y					
MC4CH	DCH Area A	Bulgaria			Y	Y	Y	Y
MolSurf	LS Area A	Bulgaria	Y					
DRS-ACS	CR Area B	FYROM	Y			Y		
CNCADD	LS Area A	FYROM	Y			Y		
BVL	DCH Area A	Romania			Y			
DREAMCLIMATE	CR Area B	Serbia	Y			Y	Y	
AUTH_WRF371M_E UR0.44	CR Area A	Greece					Y	
DACEM1	CR Area A	Greece	Y			Y		
WRF-ARW	CR Area C	Cyprus	Y					
DACEM2	CR Area A	Greece	Y			Y		

Table 1 – 1st Open Call applications and requested computational and storage services

Application Acronym	Scientific Discipline	Country	Live Access Severs	Clowder	ChemBioServer
VINE	CR Area A	Georgia			
PSOMI	LS Area A	Montenegro			
MECCA_ECHAM_GPU	CR Area A	Cyprus			
TVRegCM	CR Area A	Bulgaria	Y		
ACIQLife	CR Area B	Bulgaria			
CH-CBIR	DCH Area C	Bosnia			
Continuous_LST	CR Area A	Israel	Y		
MULTIDRUG	LS Area A	Bulgaria			
DATANMNH	DCH Area A	Bulgaria			
D3R	LS Area B	Greece			Y
MC4CH	DCH Area A	Bulgaria		Y	
MolSurf	LS Area A	Bulgaria			
DRS-ACS	CR Area B	FYROM			
CNCADD	LS Area A	FYROM			Y
BVL	DCH Area A	Romania		Y	
DREAMCLIMATE	CR Area B	Serbia			
AUTH_WRF_371M_EUR0.44	CR Area A	Greece	Y		
DACEM1	CR Area A	Greece			
WRF-ARW	CR Area C	Cyprus			
DACEM2	CR Area A	Greece			

Table 2 - 1st Open Call applications and requested application level services

4.2 Applications from the 2nd Open Call

In total 19 applications have been received in this call from 10 different countries of the region. 10 of the applications required HPC services, 9 Grid and Cloud service, 9 of storage services, 9 application specific services. The applications underwent technical review and where all the requirements were clarified they were sent for the scientific review. For confidentiality reasons the following tables show only the accepted applications.

Application Acronym	Scientific Discipline	Country	HPC	Grid	Cloud	Simple Storage	Repository	Archive	Data Analysis
DCH	DCH Area A	Israel			Y			Y	Y
mGeoAI	DCH Area C	Greece			Y				
3DGEOEX	DCH Area B	Greece			Y				
BRING-MD	LS Area A	Cyprus	Y						
OP4D	LS Area E	Bosnia and Herzegovina			Y	Y			
KPP_CLIMATE_GPU	CR Area A	Cyprus	Y						
WheAirCYEM	CR Area C	Cyprus	Y						
SIPD	LS Area A	Armenia	Y			Y			
VirMuF	DCH Area C	Egypt				Y			
3DVAR_WRF	CR Area A	Armenia	Y						
Surf_prop	LS Area A	Bulgaria	Y						
CCCSCMRR	LS Area E	Jordan							Y
OPERA-P	LS Area E	Jordan	Y	Y	Y	Y			Y
NBBM4RHMS	CR Area C	FYROM	Y			Y			
DataCrowds	DCH Area A	Cyprus							
DICOM Network	LS Area E	Moldova			Y	Y			
HaPPEn	DCH Area C	Cyprus	Y		Y	Y	Y		Y

Table 3 – 2nd Call applications and requested computational and storage services

Application Acronym	Scientific Discipline	Country	Live Access Severs	Clowder	ChemBioServer	AFMM
DCH	DCH Area A	Israel				
mGeoAI	DCH Area C	Greece		Y		
3DGEOEX	DCH Area B	Greece		Y		
BRING-MD	LS Area A	Cyprus			Y	
OP4D	LS Area E	Bosnia and Herzegovina				
KPP_CLIMATE_GPU	CR Area A	Cyprus				
WheAirCYEM	CR Area C	Cyprus				
SIPD	LS Area A	Armenia				Y
VirMuF	DCH Area C	Egypt		Y		
3DVAR_WRF	CR Area A	Armenia	Y			
Surf_prop	LS Area A	Bulgaria				
CCCSCMRR	LS Area E	Jordan				
OPERA-P	LS Area E	Jordan				
NBBM4RHMS	CR Area C	FYROM	Y			
DataCrowds	DCH Area A	Cyprus		Y		
DICOM Network	LS Area E	Moldova				
HaPPEn	DCH Area C	Cyprus		Y		

Table 4 – 2nd Call applications and requested application level services

5 Selected applications and integration efforts required

5.1 Applications from the 1st Open Call

In total 21 applications got access to VI-SEEM resources and services. The distribution of applications per country is as follows: Bosnia and Herzegovina: 1, Bulgaria: 6, Cyprus: 3, FYR of Macedonia: 2, Georgia: 1, Greece: 4, Montenegro: 1, Israel: 1, Romania: 1, Serbia: 1. During the access phase, DIOPTRA application developers changed the scope of their project and withdrew from the use of awarded resources. In total more than 14 Million CPU core hours, 3.4 Million GPU core hours, and 1 Million Phi core hours have been awarded to projects requiring among others access to the HPC services. WP6 manages the provision of resources and services to the successful projects and assigned service enablers for each one of them.

Application Acronym	Scientific Discipline	Country	HPC Resource	Grid Resource	Cloud Resource
VINE	CR Area A	Georgia	ARIS	GE-01-GRENA	
PSOMI	LS Area A	Montenegro			MK-04-FINKI_CLOUD
MECCA_ECHAM_GPU	CR Area A	Cyprus	Cy-Tera		
DIOPTRA	DCH Area B	Cyprus			CyI Cloud Facility
TVRegCM	CR Area A	Bulgaria	Avitohol		
ACIQLife	CR Area B	Bulgaria	Avitohol		
CH-CBIR	DCH Area C	Bosnia	Leo		
Continuous_LST	CR Area A	Israel			IUCC Infinity Cloud
MULTIDRUG	LS Area A	Bulgaria	Avitohol		
DATANMNH	DCH Area A	Bulgaria			Avitohol
D3R	LS Area B	Greece	ARIS		
MC4CH	DCH Area A	Bulgaria			MK-04-FINKI_CLOUD
MolSurf	LS Area A	Bulgaria	Leo		
DRS-ACS	CR Area B	FYROM	NIIFI SC		
CNCADD	LS Area A	FYROM	BA-HPC		
BVL	DCH Area A	Romania			InfraGrid
DREAMCLIMATE	CR Area B	Serbia	PARADOX		
AUTH_WRF371M_EUR0.44	CR Area A	Greece			
DACEM1	CR Area A	Greece	ARIS		
WRF-ARW	CR Area C	Cyprus	Cy-Tera		
DACEM2	CR Area A	Greece	ARIS		

Table 5 – Selected Applications and awarded computational services in the 1st Call

5.2 Applications from the 2nd Open Call

In total 17 applications got access to VI-SEEM resources and services. The distribution of applications per country is as follows: Armenia: 2, Bosnia and Herzegovina: 1, Bulgaria: 1, Cyprus: 5, Egypt: 1, FYR of Macedonia: 1, Greece: 2, Moldova: 1, Israel: 1, Jordan: 2 In total more than 2 Million CPU core hours, 1 Million GPU core hours, and 20 Thousand Phi core hours have been awarded to projects requiring among others access to the HPC services. WP6 manages the provision of resources and services to the successful projects and assigned service enablers for each one of them.

Application Acronym	Scientific Discipline	Country	HPC Resource	Grid Resource	Cloud Resource
DCH	DCH Area A	Israel			IUCC InfinityCloud
mGeoAI	DCH Area C	Greece			Okeanos
3DGEOEX	DCH Area B	Greece			Okeanos
BRING-MD	LS Area A	Cyprus	Cy-Tera		
OP4D	LS Area E	Bosnia and Herzegovina			ETFBL-CC01
KPP_CLIMATE_GPU	CR Area A	Cyprus	Cy-Tera		
WheAirCYEM	CR Area C	Cyprus	BA-HPC		
SIPD	LS Area A	Armenia	ARIS		
VirMuF	DCH Area C	Egypt			
3DVAR_WRF	CR Area A	Armenia	InfraGRID		
Surf_prop	LS Area A	Bulgaria	ARIS		
CCCSCMRR	LS Area E	Jordan			
OPERA-P	LS Area E	Jordan	Gamma	AEGIS01-IPB-SCL	MD-Cloud
NBBM4RHMS	CR Area C	FYROM	BA-HPC		
DataCrowds	DCH Area A	Cyprus			
DICOM Network	LS Area E	Moldova			MD-Cloud
HaPPEn	DCH Area C	Cyprus	Cy-Tera		Cy-Tera

Table 6 – Selected Applications and awarded computational services in the 2nd Call

5.3 Review Process

The proposals underwent a technical review and a scientific review in order to determine the eligibility and suitability of applications for the requested services and systems.

Applications requesting large amounts of HPC resources were reviewed also by independent scientific reviewers from the region taking into account conflict of interest and fairness issues. Applications not requiring HPC resources underwent a more lightweight review from scientific community leaders or other scientists. All reviews where based the criteria set in “Scope and criteria of access” of the call.

The VI-SEEM access committee comprising of the VI-SEEM project technical board prioritized the applications based on the criteria set in “Scope and criteria of access” of the call.

The applicants were notified of the final results of the evaluation. Successful applicants received further details regarding the services and resources and the process to obtain user accounts.

6 Summary of the applications

6.1 Applications from the 1st Open Call

6.1.1 Applications in Climate research

6.1.1.1 VINE

VIAM/NEA Regional Chemistry – Climate Model
Climate research – Area A

The Georgian group is concentrated on climate modelling based on the WRF climate model coupled with CHEM code for dust transportation, as dust represents main pollutant for Georgia's territory. During the past decade, there has been a significant improvement in understanding the sources, transport, properties and impacts of atmospheric dust aerosols owing to diverse new data and advancements in modelling capabilities. Georgian climate modelling group aims to improve research in process-level understanding considering the coupling and feedbacks: dust emission (influence of climate, land surface state, land use, etc.); dust ageing (cloud processing, physical and chemical interactions with other aerosols and gases); dust deposition. Significant attention is being paid to the development of a physically-based dust emission module within the Weather Research and Forecasting (WRF) model and work on assessment of dust emission parameterizations and input parameters for region.

6.1.1.2 MECCA

MECCA/ECHAM GPU Acceleration
Climate research – Area A

This application aims in using a code for accelerators for the assessment of climate change effects on pollution transport in support of Air Quality Policy formulation, pollution source apportionment and advice for impacts.

6.1.1.3 TVRegCM

Tuning and Validation of the RegCM
Climate research – Area B

Adaptation and tuning of the RegCM model for the Balkan Peninsula and Bulgaria and thus development of a methodology able to predict possible changes of the regional climate for different global climate change scenarios and their impact on spatial/temporal distribution of precipitation, hence the global water budgets, to changes of the characteristics and spatial/temporal distribution of extreme, unfavourable and catastrophic events (drought, storms, hail, floods, fires, sea waves, soil erosion, etc.). All these changes will have influence

on the ecosystems and on practically all sectors of the economy and human activity and consequently on the quality of life.

The RegCM model has many options and parameterization schemes, so the possible model configurations are more than 40. Choosing the optimal for the region of interest (Balkan Peninsula and Bulgaria) configuration by performing extensive (10 year) hindcast simulations and comparison of the results with measured data will ensure the quality and reliability of the regional climate change prediction.

6.1.1.4 ACIQLife

Atmospheric Composition Impact on Quality of Life and Human Health
Climate research – Area B

This project focusses on the development of a methodology and performing reliable, comprehensive and detailed studies of the impact of lower atmosphere parameters and characteristics on the quality of life (QL) and health risks (HR) for the population in our country.

Thus formulated objectives contain several key words, which have to be explained: Methodology: this is the totality of metrics for evaluation of the atmospheric parameters impact on the quality of life and health risks for the population; a set of properly chosen and well verified and validated models of atmospheric dynamics and chemical composition; data bases (including data for the health status (HS) of the population) and procedures for access to data and data processing; a set of appropriately defined scenarios for extensive computer simulation experiments; appropriate statistical instruments and strategy for analysis and generalization of the data for the population HS.

Reliable and comprehensive studies: This means carrying out of extensive and appropriately enough defined numerical experiments, which to form statistically significant ensembles of output data, which reflex the diversity of meteorological conditions with their typical recurrence and which, combined with the available HS data, to allow making reliable conclusions for the atmospheric characteristics impact on population quality of life and health risks.

Detailed studies: This means high enough spatial/temporal resolution of the computer simulations, which to reflect the multi-scale nature of the processes, to make it possible the detection of interactions of different scale phenomena and tracking the basic mechanisms and pathways trough which low atmosphere characteristics are formed, respectively their impact on population quality of life and health risks.

It is envisaged the studies to be performed for the whole country, but emphasizing on urban environment studies, because the specific climatology and ecology of urban environment make it less favourable from a point of view of population quality of life and health risks. The totality of meteorological fields and AQ of the lower atmosphere will be further generally

called Air Environment (AE) and Urban Air Environment (UAE)." Evaluation and analysis of the AE and UAE characteristics impact on the QL and HH – revealing the connections between the parameters of AE and UAE in Sofia with the QL and HH; defining the sets of parameters that determine the adverse effects and critical consequences for the QL and HH; specify factors (primarily a large-scale meteorological conditions and emissions from different types of human activity) that influence the formation of adverse and critical parameters of AE and UAE; identifying of strategies and practical measures for adaptation and/or mitigation of the adverse impact of the AE and UAE on QL and HH. Multi-scale computer simulations, generating of multi-scenarios ensemble of computer simulations: The study of AE and UAE will be based on synergistic use of measured data and numerical simulations. Numerical simulation is a fruitful approach, which helps better understanding the part for which different processes and transport scales play in AE and UEA formation. The cohesion of models and data is of crucial importance, so the proper (also regarding the considered spatial/temporal scales) choice of approaches, methods, tools and modelling practices, as well as reaching good experience and operational skills in using these tools are as important for AE/UAE studies as the input data.

6.1.1.5 Continuous_LST

Continuous Land Surface Temperature
Climate research – Area A

Land Surface Temperature (LST) is a key variable in climatological and environmental studies. It can be used for monitoring vegetation water stress, assessing surface energy balance, detecting land surface disturbance, and public health (exposure assessment and monitoring condition suitable for vector-borne disease). Hence, a global continuous LST is a vital product. Satellite LST products provide an estimate of the kinetic temperature of the earth’s surface skin which can be determined from thermal emission at wavelengths in the thermal infrared. However, these measurements are limited to clear sky conditions. The polar-orbiting MODIS sensors, produce daily LST maps with global coverage. Climatology of cloud free LST was generated for time series (2002-2014) of MODIS AVHRR over TERRA data. From this time series one can learn the seasonal behaviour of each pixel (climatology). Actual LST depends on both the climatology (seasonal mean) and the anomaly attributed to the weather determined by synoptic scale circulation.

In order to calculate the temperature deviation from normal seasonality temperatures (anomaly), Numerical Weather Prediction (NWP) data and NCEP Climate Forecast System Version 2 (CFSV2) are used. NWP models gives a global temperature prediction regardless of the cloud cover. MODIS LST is at high spatial resolution (1 km), while surface temperature from the model is at lower resolution (25 km).

By combining high resolution climatological LST from satellite with the synoptic influence from the model the actual LST values of cloudy pixels can be estimated.

6.1.1.6 DRS-ACS

Dynamics, reactivity and spectroscopy of atmospheric chemical species
Climate research – Area B

Novel hybrid methodologies, based on statistical physics (SP) and quantum mechanics (QM), will be developed to model the dynamics, spectroscopic properties and reactivity of individual molecular species and noncovalently bonded molecular clusters relevant to atmospheric chemistry under realistic conditions encountered in the Earth's atmosphere (finite temperatures, irradiation etc.). The statistical physics sequence will involve either classical or quantum (i.e. ab initio) Monte Carlo (MC) or molecular dynamics (MD) simulation of the molecules/clusters under the appropriate conditions. Ab initio MD methods will be based on Born-Oppenheimer approach (BOMD), the Car-Parrinello approach (CPMD), or using the atom-centered density matrix propagation scheme (ADMP). A series of statistically uncorrelated configurations ("snapshots") from MC or MD simulations is picked up from the equilibrated SP phase runs and subjected to the sequential QM computational phase. The choice of statistically independent configurations is based on time-series analytic methods (i.e. analysis of appropriate time correlation functions). Appropriately chosen configurations from the SP runs are further treated by the state of the art QM methodologies (Moller-Plesset perturbation theory, coupled cluster and configuration interaction methods, density functional theory, including time-dependent DFT etc.), in the sense of computation of certain structural and spectroscopic properties relevant to their detection, as well as reactivity. During this phase, the main accent is being put on the modelling of energetics of the interactions, anharmonic vibrational frequencies (including the calculations of the complete bands in the infrared or Raman spectra), magnetic characteristics (e.g. isotropic shielding constants, coupling constants etc.), as well as some other structural and dynamical characteristics of the studied systems. Computation of all these properties is carried out in a most rigorous and exact manner. For example, to compute the relevant anharmonic vibrational frequencies of particular intra- and intermolecular modes, which are directly comparable to the experimental vibrational spectroscopic data, appropriate cuts through the vibrational PES corresponding to the normal (or local) modes in questions are computed in a pointwise manner. Either 1D, or in some cases 2D, 3D or even nD vibrational potentials are computed and the corresponding 1D, 2D, 3D or nD vibrational Schrodinger equation is solved. Further, to convert the calculated vibrational density of states histograms (which are directly comparable to the Raman spectral bands) into scaled histograms (due to the dependence of the transition dipole moment on the frequency) which are directly comparable to the infrared spectral bands, an exact ab initio methodology is used (i.e. usage of any empirical scaling factors is avoided). In such manner, the influence of inter- and intramolecular dynamics on molecular structure and properties of species relevant to atmospheric chemistry will be explicitly and rigorously accounted for. At the same time, these data will substantially improve the in-depth understanding of the physics and chemistry of a wide variety of species and processes taking place in Earth's atmosphere, related to air pollution, ozone layer degradation and numerous other issues.

6.1.1.7 DREAMCLIMATE

Dust Regional Atmospheric Model Climatology
Climate research – Area B

The top-level objective of this project is the utilization of the gridded dust model. This is relevant for various impact studies such as: potential dust effects on the human mortality in the North Africa – Europe – Middle East region, exploration of possible links between dusty weather and meningitis epidemics in Africa, cloud-dust aerosol interactions, model validation against remote sensing observation (e.g. lidar observations), etc.

The gridded dust model outputs will be produced by software developed and implemented within the group - Dust REagional Atmospheric Model (DREAM). This model provides a climatology of dust based on the long-term re-analysis. It is widely used (in more than 20 countries) in the research and operational dust forecasting communities, including its recent use in a series of NASA-funded projects dealing with health aspects of dust suspended in the air. WMO in SDS-WAS project (http://sds-was.aemet.es/), uses daily dust forecasts produced by DREAM for the SDS-WAS model inter-comparisons and validation. Within the project a dataset with one-year modelling climatology for the region is being generated. The dataset is produced by running the DREAM model in a high-resolution mode with the horizontal grid resolution of about 5-10 km.

6.1.1.8 DACEM1

Dust Aerosols Climatic Effects over the Mediterranean - Phase 1
Climate research – Area A

There is currently solid evidence that dust emitted from arid and desert regions around the globe influences significantly the climate system. However, there is an ongoing debate with respect to the dust aerosols effects and associated feedbacks that influence precipitation formation, and cloud properties and lifetimes. As a result, modelling of dust and its radiative effects and interactions with clouds and precipitation remains a daunting task. A large part of the uncertainty in simulating the climatic effects of dust arises from the estimation of size-resolved emissions in the source regions, the treatment of aerosols in numerical models, and the determination of optical properties of dust. Nevertheless, it has been shown that numerical weather prediction (NWP) models are capable of properly representing the radiation budget when a proper treatment of dust and its radiative effects is employed. Ultimately, more studies are needed on the dust-cloud-precipitation interactions to gain a useful insight for model diagnosis and identify improvements that are critical for projecting future climate change.

Within the above content, this project aims at quantifying the long-term climatic effects of dust aerosols over the Mediterranean basin, a region of the globe that has been well documented as a climate hot-spot. A state-of-the-art coupled meteorology-chemistry model, namely WRF-Chem, is employed in order to examine the effects desert dust particles exert

on the climate system. For this, decadal simulations, with and without the inclusion of dust emissions, will be carried out at high horizontal grid spacing.

6.1.1.9 AUTH_WRF371M_EUR0.44

AUTH contribution to EURO-CORDEX 0.44 regional climate projections
Climate research – Area A

The provided regional climate dataset is an ensemble member produced by the Department of Meteorology and Climatology, School of Geology and Aristotle University of Thessaloniki, in the framework of EURO-CORDEX. The CORDEX (Coordinated Regional Downscaling Experiment) initiative operates under the auspices of the WCRP (World Climate Research Project) and aims to produce coordinated sets of regional downscaled projections worldwide. The applicants use the Weather Research and Forecasting (WRF) model (v3.7.1) to provide a hindcast simulation extending from 1990 to 2008, with 3hourly temporal analysis covering the European domain with 50 Km spatial resolution forced by the ECMWF-ERA-interim reanalysis. The same configuration of WRF371 is used to produce a regional climate projection driven by the NASA-GISS_E2 global circulation model for the years 1975 to 2099 under the RCP8.5 scenario.

6.1.1.10 WRF-ARW

WRF-ARW: Weather forecasting with chemistry over Cyprus
Climate research – Area C

Provision of user-friendly, detailed weather forecasts to the general public, including warnings for severe weather in order to prevent/minimize the impacts of such weather to daily activities. Investigation of the impact of pollution on weather parameters.

6.1.1.11 DACEM2

Dust Aerosols Climatic Effects over the Mediterranean - Phase 2
Climate research – Area A

There is currently solid evidence that dust emitted from arid and desert regions around the globe influences significantly the climate system. However, there is an ongoing debate with respect to the dust aerosols effects and associated feedbacks that influence precipitation formation, and cloud properties and lifetimes. As a result, modelling of dust and its radiative effects and interactions with clouds and precipitation remains a daunting task. A large part of the uncertainty in simulating the climatic effects of dust arises from the estimation of size-resolved emissions in the source regions, the treatment of aerosols in numerical models, and the determination of optical properties of dust. Nevertheless, it has been shown that numerical weather prediction (NWP) models are capable of properly representing the radiation budget when a proper treatment of dust and its radiative effects is employed. Ultimately, we need more studies on the dust-cloud-precipitation interactions to gain a useful insight for model diagnosis and identify improvements that are critical for projecting future climate change.

Within the above content, this project aims at quantifying the long-term climatic effects of dust aerosols over the Mediterranean basin, a region of the globe that has been well documented as a climate hot-spot. A state-of-the-art coupled meteorology-chemistry model, namely WRF-Chem, is employed in order to examine the effects desert dust particles exert on the climate system. For this, decadal simulations, with and without the inclusion of dust emissions, are carried out at high horizontal grid spacing.

6.1.2 Applications in Life Sciences

6.1.2.1 PSOMI

Protein-Small Organic Molecules Interaction
Life Sciences - Area A

The basic of every bio-activity is ligand-receptor interaction. In order to have better understanding of bio-activity of whole tested organic molecule group, the molecular dynamics simulation approach is used. Testing the interaction of the (2R, 3R, 4R, 5S) -5-Hydroxy-5 -((S) -methylene-2-oxo-1, 3-dioxolan-4-yl) -pentan-1,1,3,4-tetraacetate and similar molecules with protein provides a way to find and analyse the binding and active site of the protein. With this approach it is possible to design structurally related molecules that could have biological and physiological effects. In the absence of biological tests, molecular dynamics method is certainly one of the best choices for such testing.

Lack of pharmaceutical industry and medical research as well as lack of funds for such research and testing on live objects is main reason for such situation. The MD simulation methodology which is applied for the first time allowing to overcome those obstacles and solve the problem of testing on bioactivity.

This testing model could provide the opportunity to test various organic compounds and radicals in different concentration, and determine if they have some negative effects on live components of ecosystems. Hence, we can even do the fine calibration of present Maximal Allowed Concentration of different pollutants that are mainly residuals of organic compounds.

6.1.2.2 MULTIDRUG

Multiscale computations to reveal the potential of molecular drug carriers
Life Sciences – Area A

The main disadvantage of current drug delivery systems (DDS) is that the active substance is distributed in a non-specific manner. Doxorubicin (DOX), which is one of the most widely used chemotherapeutics for cancer cure, is a good example, which will be used as a model drug. It is a representative of the anthracycline antibiotics and is suitable for treatment of solid tumours, leukaemia, lymphomas, or soft tissue sarcomas. However, the success of DOX is limited by the severe adverse effects that follow its administration. To improve the efficiency

of its action, multi-block molecular carriers are needed. The DDS has to comprise driving vehicles, i.e., entities that bind multiple drug molecules and can carry them as cargo in vivo, targeting agents to steer the drug to specific tissues or cell compartments, and parts that facilitate the passage of the active substance across cell membranes. A multitude of compounds and assemblies have been tested as building blocks of such transporting moieties but none are widely acknowledged as the most suitable ones. In the present project, optimization of the targeting step of the process will be in focus. Various vector ligands that facilitate recognition by important malignant cell membrane receptors are screened and the most prospective one(s) are attached to a drug-vehicle system to study the mechanism of receptor recognition and of penetration of the model DDS across cell membrane models. This aids optimization of the composition of such multi-block drug delivery systems with respect to function. The processes will be studied at the molecular level by atomistic and coarse-grained classical molecular dynamics simulations of model DDS (and model cell membrane) in their explicit natural aqueous environment.

6.1.2.3 D3R

Predicting Farnesoid X receptor (FXR) - inhibitor structures and affinities for computer-aided drug design within the D3R challenge
Life Sciences – Area B

Computational tools, such as docking and scoring, for modelling the interactions of proteins with drug-like compounds (ligands) hold great promise to speed the discovery of new, safer medications and reduce the cost of the drug discovery process. However, there is still a need for improved methods of predicting ligand poses and affinities or relative affinities.
In this project, the investigators were provided by a blinded unpublished dataset, within the D3R challenge, containing high quality crystal structures and potency data for testing and improving ligand-protein docking algorithms and their scoring protocols. It is based on the Farnesoid X receptor (FXR) target; the dataset is kindly contributed by Roche and curated by D3R.

The FXR nuclear receptor forms a heterodimer with RXR when activated, and binds to hormone response elements on DNA, leading to up- or down-regulation of the expression of certain genes. FXR agonists are regarded as potential therapeutics for dyslipidemia and diabetes.

The blind dataset has 36 crystal structures with resolution < 2.6Å; and binding data (IC50s) for 102 compounds across five orders of magnitude (Scintillation Proximity Assay), comprising four chemical series and 6 miscellaneous compounds. The researchers have been provided only with the structures of the 102 ligands and using publicly available data about FXR are called to: a) Predict the crystallographic poses of 36 ligands spanning all chemical series, b) Predict affinities, or affinity rankings, for these ligands, and also for the other 66 ligands, c) Predict the relative binding affinities for two designated free energy subsets of 18 and 15 compounds.

This project aims to help advance computational technologies such as docking and scoring and free energy methods by using unknown ligand-protein datasets, predicting their outcome and helping advance the state of the art of the computer-aided drug design techniques.

6.1.2.4 MolSurf

Molecular description of the rheological properties of natural and synthetic surfactants at the air-water interface.
Life Sciences – Area A

Classical molecular dynamics simulations are employed to monitor the behaviour of two synthetic (LAS (here LAS stands for Linear Alkylate Sulfonate) and SLES) and natural (Escin) anionic surfactants at the water/vacuum interface. In the case of the natural surfactant Escin which has a noticeable bioactivity, it demonstrated self-assembly behaviour, as a result of which the molecules were well organized via dispersion forces and a significant number of hydrogen bonds. These two types of interaction are the most probable causes for the unique rheological characteristics of saponin adsorption layers. In the case of synthetic surfactants, LAS and SLES show similar behaviour when there are no Ca2+ ions in the solution but when Ca2+ is added Na+ is completely displaced from the surface with LAS molecules and partially from the surface with SLES molecules. This behaviour is explained by the specific interaction between the benzene ring in the LAS molecule, which is absent in the system with SLES and Ca2+. All these predictions have been made from MD simulations of relatively small models. To confirm the hypotheses, it is necessary to carry out molecular dynamics simulations on much larger models mimicking an actual adsorption layer, which will verify whether these are the actual causes of the experimentally observed non-trivial rheological behaviour of Escin and of the selective binding of Ca2+ to LAS head groups. Furthermore, the more extensive computing power will allow investigation of the aggregation and adsorption processes. Elucidation of the origin of the unique characteristics of the surfactants at molecular level will have both a fundamental and a practical impact.

6.1.2.5 CNCADD

Conventional vs Novel Computer Assisted Drug Delivery
Life Sciences – Area A

Within the current project, a multi-disciplinary scientific approach will be developed, with ICT providing envelope for extensive research activities involving and integrating achievements and advancements in several scientific domains such as computational biology, chemistry and pharmaceutics, with a primary focus on modelling of the intermolecular interactions relevant to the design and actual development of efficient smart drug delivery nanocarriers (SDDN) specifically aimed for cancer treatment. Novel sequential methodologies, based on combination of statistical physics (SP) and quantum mechanics (QM) approaches are being developed to model the dynamics, spectroscopic properties and reactivity of individual molecular species and noncovalently bonded molecular clusters relevant to design and fabrication of SDDN. Several fundamental questions will be addressed: i) How does the spatial

confinement affect the type and strength of biologically relevant intermolecular interactions; ii) How does the spatial confinement affect the reactivity of biologically relevant molecular systems: in this context, special attention will be paid to simple reactions such as e.g. proton-transfer in tautomeric equilibria, or the processes of transfer of whole polyatomic groups; iii) How does the spatial confinement affect some fundamental molecular properties, such as e.g. acidity or basicity of particular molecular centres. Under the condition of nanoconfinement, one deals with a fluctuating nanofluidic environment which needs to be carefully accounted for in order to obtain a realistic theoretical model of the physical situation encountered in biological nanofluids. The fluctuating character of the nanoenvironment is accounted for applying sequential statistical physics – quantum mechanical approach. The statistical physics sequence involves either classical or quantum (i.e. ab initio) Monte Carlo (MC) or molecular dynamics (MD) simulation of the molecules/clusters under the appropriate conditions. Ab initio MD methods are based on Born-Oppenheimer approach (BOMD), the Car-Parrinello approach (CPMD), or using the atom-centered density matrix propagation scheme (ADMP). A series of statistically uncorrelated configurations ("snapshots") from MC or MD simulations are picked up from the equilibrated SP phase runs and subjected to the sequential QM computational phase. The choice of statistically independent configurations is based on time-series analytic methods (i.e. analysis of appropriate time correlation functions). Appropriately chosen configurations from the SP runs are further treated by the state of the art QM methodologies (Moller-Plesset perturbation theory, coupled cluster and configuration interaction methods, density functional theory, including time-dependent DFT etc.), in the sense of computation of certain structural and spectroscopic properties relevant to their detection, as well as reactivity. During this phase, the main accent is being put on the modelling of energetics of the interactions, anharmonic vibrational frequencies, magnetic characteristics, as well as some other structural and dynamical characteristics of the studied systems. Computation of all these properties is carried out in a most rigorous and exact manner. For example, to compute the relevant anharmonic vibrational frequencies of intra- or intermolecular modes, which are directly comparable to the experimental vibrational spectroscopic data, the corresponding 1D, 2D, 3D or nD vibrational Schrodinger equation needs be solved.

6.1.3 Applications in Digital Cultural Heritage

6.1.3.1 CH-CBIR

Content-Based Image Retrieval and Classification in Cultural Heritage Applications Digital Cultural Heritage – Area C

Convolutional neural networks (convnets) have shown excellent results in various image classification tasks and remote sensing image classification is not an exception. However, successful end-to-end training of deep convnets requires large amounts of labeled training data. Unfortunately, in remote sensing applications data labelling is expensive and large labelled datasets are scarce. On the other hand, a number of research studies have shown that convnets trained on one classification problem can be used on a different task with minor modifications and without the need for time-consuming retraining. More specifically,

previously trained (pretrained) convnets can be used as feature extractors and the obtained features can be used for image retrieval or classification regardless of the specific classifier.

The most commonly encountered pretrained convnets are trained on ImageNet dataset which contains 1.2 million images which are manually classified into 1000 categories of objects. However, the distribution of remote sensing images is not completely the same as the distribution of ImageNet images. Namely, since ImageNet mostly contains images of objects, texture does not play an important role in their classification. On the other hand, texture is an important cue in remote sensing image classification. Moreover, images in ImageNet are obtained using the visible part of the spectrum only, while remote sensing images are often multispectral. In this case the learned filters are not able to use the information from the additional spectral bands. The consequence of these findings is that there is still room for improvement of convnets when the task at hand is remote sensing image classification.

Recently, new large datasets of remote sensing images have been made available. The size of these datasets enables successful training of convnets. Therefore, it is possible to try and use the convnets trained on these datasets for feature extraction in other remote sensing image retrieval and classification problems. Since the distributions of images in the source and target problems are now much more similar we expect that the performances of the resulting systems will be better than in the case when ImageNet classification is used as a source problem.

Nevertheless, significant variations are still possible between the images in the source dataset and images in a target dataset. These variations are mainly due to the differences in the used sensors, imaging conditions, data postprocessing, etc. The goal of this project is to investigate the generality of features extracted using pretrained convnets in tasks of remote sensing image retrieval and classification. To this end we consider convnets from the perspective of feature extraction for remote sensing image classification. We also analyse the approaches to using the features extracted using a convnet for image retrieval and classification in other remote sensing datasets. As an outcome of this project we propose an approach for extraction of discriminative features from remote sensing images using pretrained convnets. Since training of convnets on large datasets is time and resource consuming it is our intention to make publicly available trained models which can be used for feature extraction in other datasets.

6.1.3.2 DATANMNH

Databases about National Museum of Natural History Collection
Digital Cultural Heritage – Area A

The Bulgarian National Museum of Natural History is the only Bulgarian institution directly engaged with the preservation of scientific collections of live and non-live nature from Bulgaria and the world. The study of biodiversity, environmental protection and the evolution of organisms are the museum's major priorities. Consequently its main task is the all-around study of the fauna, flora, fossils, minerals and rocks of Bulgaria and other countries. With this

project proposal, the applicants aim to digitize the information about the National Museum of Natural History collections in order to reach more effectively public and scientific visitors.
The Museum activities related to maintenance the exposition, its enrichment and the processing of scientific collections are performed along with the scientific studies. By using the VI-SEEM project services, it is possible to develop a very important web tool for the museum visitors to have access to the biodiversity data, for educational and scientific purpose. This makes more accessible the information and helps to preserve the collections' specimens.

6.1.3.3 MC4CH

Management Cloud for Cultural Heritage
Digital Cultural Heritage – Area A

The aim of the project is to propose a geospatially-enabled software platform for cultural heritage inventory and management.

The open-source system Arches is set up and managed for this purpose. A number of pilot applications will be presented.

6.1.3.4 BVL

Banatica Virtual Library
Digital Cultural Heritage – Area A

The BANATICA collection gathers together all the printed products considered monographs (brochures, books, yearbooks, calendars in volumes, prints with an individual cover, atlases, book-like printed scores etc.) which represent documentation sources for the culture and civilization on the Banat region. This collection was jointly created by the Central University Library "Eugen Todoran" Timisoara (Romania) and "Zarko Zrenjanin” public library (Serbia) in order to wide-spread the culture and history of Banat region, a geographical and historical region in Central Europe that is currently divided among three countries: the eastern part lies in western Romania, the western part in northeastern Serbia and a small northern part lies within southeastern Hungary. The collection contains 1000+ bibliographic descriptions and 200 full-text scanned books.

The Banatica Virtual Library (BVL) project aims to establish a virtual database of patrimonial books based on this existing collection through the following objectives:

- Make it available to a wider, international community within the region.
- Make it machine-searchable and indexable.
- Make it machine readable.

The first objective -- make it available to a wider, international community -- will be addressed by sharing the collection through Clowder service provided by VI-SEEM Virtual Research Environment. In order to make the collection machine-searchable and indexable, the

bibliographic descriptions, including title (translated in four languages), authors, place of publication, language of publication, will be integrated within Clowder's search engine.

The third objective -- make the collection machine readable, requires applying OCR techniques on digitized books. Considering the age of the collection, applying OCR is not a trivial task:

- Being +150 years old, the quality of paper prints is poor thus scanned images are noisy.
- The monographs contain end-users notes and other undesired artifacts.
- The vocabulary used in these books differs from current vocabulary.
- Mixed character sets used, Latin and Cyrillic.
- In case of Latin text, there are specific diacritics used that are not in used anymore in that language.
- Scanned material consists of books between tens up to 200 pages.

Due to these challenges, existing OCR tools need to be adapted, customized to cope with poor quality images, recognize old word forms, recognize old diacritics etc.

6.2 Applications for the 2nd Open Call

6.2.1 Applications in Climate

6.2.1.1 KPP_CLIMATE_GPU

KPP / KP4 Climate GPU Acceleration
Climate research – Area A

The global climate model ECHAM/MESSy Atmospheric Chemistry (EMAC) is used to study climate change and air quality scenarios. The EMAC model is constituted by a nonlocal dynamical part with low scalability, and local physical/chemical processes with high scalability. The EMAC chemistry--climate model does not benefit from the support of accelerators which are nowadays installed in many HPC systems. We will study strategies to offload the calculation of the atmospheric chemistry to accelerator technologies (GPU and Intel MIC), as in typical model configurations this is the most computational resource-demanding subtask. The proposed solutions extend the Kinetic Pre Processor (KPP) general purpose open-source software tool used in atmospheric chemistry.

The researchers will develop and optimize the Rosenbrock family of solvers for ordinary differential equations on accelerator (GPU) architectures. The proposed solutions extend the Kinetic Pre Processor (KPP) general purpose open-source software tool used in atmospheric chemistry.

6.2.1.2 WheAirCYEM

Weather and air pollution forecasting for Cyprus and Eastern Mediterranean
Climate research – Area C

The aim of the project is to provide air quality/chemical weather forecasts, for the region of Eastern Mediterranean, with a focus on Cyprus. The forecasts will include extreme event (e.g., dust storms) warnings in order to help mitigate the impacts of such weather phenomena to daily activities. The researchers will further explore the impact/feedback of air pollution patterns on general weather parameters.

The research project will include the following:

1. Meteorological and air pollution forecasting for Eastern Mediterranean and Cyprus. For the latter the simulations will be performed on a very high-resolution grid.
2. Impact of chemical species on weather patterns.
3. Forecast of extreme events such as dust storms.
4. Study the links and feedbacks between aerosols and cloud, radiation and precipitation.

6.2.1.3 3DVAR_WRF

Using WRF-3DVAR assimiliation system to improve the weather prediction over the Ararat Valley Region of Armenia
Climate research – Area A

An important factor in the accuracy of forecasts produced by numerical weather prediction models is how well the model`s initial conditions match the actual conditions. In general, more accurate initial conditions should lead to a more accurate forecast. In this study we use the WRF model, which has a disadvantage of low accuracy in weather prediction. One reason of low accuracy of the model is in accuracy of initial condition model to the actual atmospheric conditions. Techniques to improve the initial condition model is the observation data assimilation. This study is devoted to the evaluation of the role of assimilation of conventional data on temperature, humidity forecasts at a regional scale. The analysis based on time scale modeling from the heat season that occurred on the territory of Armenia's limited Ararat valley. Observational data used in data assimilation are observation data from all stations. The aim of the study is to compare the effect of data assimilation with an additional data observations obtained from the weather stations.

6.2.1.4 NBBM4RHMS

NBBM Modeling for RHMS
Climate research – Area C

Weather significantly affects the health, safety and economic well-being of all. Climate change and global warming contribute to the more frequent occurrence of weather and hydrological

disasters that have intensified intensity and adversely affect people and material goods, and hence to sustainable development. That is why continuing progress with respect to the methods and tools for forecasting the time, monitoring new technologies in development and applying the numerical forecast of time as a starting point for the preparation of timely precise and detailed weather information is of utmost importance.

In order to address these climate change challenges, the Republic of Northern Macedonia, as a member of the newly established SEECOP Operational Numerical Forecast Consortium, with the commitment of the National Hydrometeorological Service and cooperation with FINKI UKIM, is accelerating its efforts towards improving resources in the area of climate change. The preparation of the establishment and implementation of an integral contemporary prognostic system for time, seasonal and interim forecasts as well as the preparation of climate scenarios future climate in Macedonia. This will be achieved through an integrated approach, using regional climate models or available tools, with the sole aim of strengthening the use of climate information and forecasts for decision-makers in more socio-economic sectors.

6.2.2 Applications in Life Sciences

6.2.2.1 BRING-MD

Large scale molecular dynamics simulations of BRCA1-RING domain missense mutations
Life Sciences – Area A

Germline mutations in BRCA1 gene are associated with an increased lifetime risk for developing breast and/or ovarian cancer. Currently, 1458 BRCA1 missense mutations have been reported to the ClinVar archive. Of these, 79 have been confirmed as being pathogenic, 150 as being bening and over 1200 are characterized as Variants of Uncertain clinical Significance (VUS) i.e. variants whose pathogenic or benign effect has not been demonstrated. More than 50% of the pathogenic missense substitutions occur within the N-terminal RING domain [1], which is responsible for the E3-ubiquin ligase activity of BRCA1 [2]. Up to date, 62 missense mutations have been identified in the RING domain. Of these, 19 have been classified as pathogenic, 3 as benign and 40 are characterized as VUS. It is known that the BRCA1 RING domain forms a heterodimer complex with the RING domain of BARD1. The formation of this heterodimer complex has been shown to increase dramatically the E3 ubiquitin ligase activity of BRCA1 [3]. Hence, the stability and thus functionality of BRCA1 protein is related to the structural stability of the aforementioned heterodimer. Indeed, functional studies of BRCA1 variants demonstrated that pathogenic missense substitutions on the BRCA1 RING domain were also affecting the BRCA1/BARD1 heterodimerization. Furthermore, ubiquitination, which is the ligation of a ubiquitin moiety or a polyubiquitin chain to protein substrates, is mediated by interaction of an E3-ligase enzyme (such as the BRCA1 RING domain) with an E2-ubiquitin conjugating enzyme. The BRCA1 RING domain binds to E2 enzymes through a surface that is opposite the binding interface with BARD1 [4]. The vital

role of BRCA1 E3 ligase activity is demonstrated by its targets, which include estrogen receptor-alpha, progesterone receptor, CtIP and histone protein H2A. Therefore, missense BRCA1 RING domain missense mutations that disturb the BRCA1–BARD1 heterodimer complex or BRCA1-E2 interaction, may alter BRCA1 function and lead to an increased risk of cancer. For that reason, investigation of the effects of BRCA1 RING domain missense mutation upon the stability of the aforementioned complexes will contribute significantly in gaining insights into molecular mechanisms underlying the involvement of BRCA1 in tumorigenesis. However, traditional functional assays require considerable time, effort and skills and applying such demanding approaches to study each BRCA1 variant is infeasible. Here, we suggest the implementation of molecular dynamics (MD) simulations in order to investigate the influences of BRCA1 RING domain missense substitutions upon the structure of BRCA1 RING domain as well as upon the formation of the BRCA1–BARD1 RING domain heterodimer and BRCA1–E2 complexes. In addition, it has been suggested that inhibition of BRCA1-BARD1 RING heterodimer may sensitize tumors to PARP inhibition. Using data generated by MD simulations and virtual screening, we aim to identify small molecules that could selectively inhibit BRCA1-BARD1 heterodimer and or BRCA1-E2 interaction, where BRCA1-RING domain variants are involved.

6.2.2.2 OP4D

OrthoPhoto4D
Life Sciences – Area E

In orthodontic diagnostic process, the orthodontist needs to perform a series of measurements, traditionally by manual process using callipers. Recently there is a movement to digitize study models (plaster casts of teeth) and then perform measurements. Although the scanners have come down in price recently, they are still quite expensive and out of reach for most, especially those that are precise enough for this use. This is why we are developing a method to use photogrammetry in measurements. Operator takes 4 photographs using marked device for holding measured object and then processes the photographs in the program that is also used for measurements in 3D space based on four 2D photographs. Device is designed to accomodate relatively small objects (approx 80x65x40mm). The software also preforms two analysis used in orthodontics (Bolton and Lundstromm). Also a web based software for measuring 3D objects (generic) is being co-developed with existing DCH efforts related to digitalization of Museum of RS collection.

6.2.2.3 SIPD

Surfactant induced protein denaturation
Life Sciences – Area A

The computational simulations (mainly molecular dynamics simulations) are of special relevance in bio-modeling, and pharmaceutical applications, e.g. as potential drug delivery systems. This kind of computational experiments offers a detailed picture of the structure and

dynamics in the multicomponent system and helps therefore to gain a better understanding of the mechanism of interactions between compounds, as a result, help us to modulate exchange processes between them. Surfactant/protein complexes have been intensively examined as their applications in many fields from medicine to oil industry. These systems are of practical interest due to the ability of surfactant to denature/renature proteins. A long timescale simulation is planned to check influence of sodium dodecyl sulfate (SDS) on protein aimed to elucidate how the SDS operates at molecular level and reveal the mechanism of protein folding in terms of atomic interactions. In order to get a more realistic model close to natural systems, it is necessary to simulate systems of ns-to-ms in length. Therefore, a longer simulation time is needed, and the application of parallel powerful computers availability of high performance facility and supermachine resources has become very useful.

6.2.2.4 Surf_prop

Physicochemical characterization of the natural and synthetic surfactant systems with unique properties caused by specific interactions
Life Sciences – Area A

Classical molecular dynamics simulations are employed to monitor the behaviour of synthetic (LAS and SLES) and natural (Escin) anionic surfactants in the bulk and at the water/vacuum interface, respectively. In the case of the natural surfactant Escin, it demonstrated self-assembly behaviour on the surface, as a result of which the molecules were well organized via different types of interactions. The main intermolecular forces governing the escin self-assembly are: (1) long-range attraction due to the inhomogeneous charge distribution in the aglycone and the dispersion (London) van der Waals forces between the aglycone fragments, (2) intermediate-range dipole-dipole interaction, (3) shortrange classical H-bonds, and (4) electrostatic repulsion between the charged carboxyl groups in the ionized escin molecules. These types of interaction are the most probable causes for the unique rheological characteristics of saponin adsorption layers. The objective of the current project is to check whether the types and balance of interactions will be preserved also in condensed adsorption layers.

In our previous studies on the behaviour of synthetic surfactants, specific interaction between the benzene ring in the LAS molecule and Ca2+ was detected, which is absent in the system with SLES. The experimental data about surface area per molecule was reproduced by monitoring of the adsorption process. To investigate the effect of electrolyte, which is observed in the experiment, it is necessary to simulate model systems in the presence of electrolyte. To explain the experimentally observed phenomena, it is reasonable to carry out molecular dynamics simulations on models mimicking the experimental conditions, which will verify the mechanisms of the specific interactions that lead to experimentally observed non-trivial behaviour of the studied molecules. Furthermore, the more extensive computing power will allow investigation of more realistic molecular systems. Elucidation of the origin of the unique characteristics of the surfactants at molecular level will have both a fundamental and a practical impact.

6.2.2.5 CCCSCMRR

Using the CCC software to compare the result with mammography radiologist results
Life Sciences – Area E

The aim for this project is to test the ability of computer data processing for image in order to detect abnormalities (mainly the cancer cells), and could detect hidden lesions that could not be detected by radiologist eyes follow up of the patient to see the result. The software used is CCC developed by Synchrotron-light for Experimental Science and Applications in the Middle East (SESAME).

6.2.2.6 OPERA-P

OPportunistically, Elastically Resource Allocation and Provisioning task scheduler
Life Sciences – Area E

Handling the increasingly and unprecedented amount of data sets demands a new scalable and extendable analytical platforms. Nowadays, Big Data (BD) Analytics and Large-Scale Distributed Systems (LSDS) are the key research subject for accommodating the varying needs of these applications. However, there is a notable increasing gap between computation and I/O capacity on high-end workstations, makes a severe bottleneck for BD operations. Scalability and elasticity of distributed systems (Cloud computing or on-premise) will be essential for next generation of BD architectures. To enhance system throughput, current platforms and paradigms, typically involve statically adding new expensive nodes. Also, current approaches do not consider optimal utilization of valuable resources within a physical confine of an enterprise in a dynamic manner. To address these issues, we propose OPERA-P, short for OPportunistically, Elastically Resource Allocation and Provisioning scheduler, a new hybrid analytical environment that integrates dedicated and nondedicated nodes in one framework.

OPERA-P deploys an Opportunistic Container-based Cluster (OCBC) as a new Container-as-aService (CaaS) over the resources pool, which dynamically extends the cluster capabilities on demand. Using an adaptive task scheduling algorithm, where volunteer nodes are used to supplement the logical servers, i.e., by deploying pilot containers, they execute tasks and then die. This Dynamic Provisioning On-demand (POD) approach, exploits commodity personal workstations, in a controlled network, e.g., enterprise desktop Grid. Thus, not only enables sharing all resources, including data, across different workloads to improve time-to-value aspects but also ensures minimizing the cost of deployment, and reduces operational costs. It reduces the total cost of ownership and gaining a better return on infrastructure investment as well. EXTRA INFO FROM PI: BOINC-like'' describes well but partially the work that has been done on my platform OPERA, which stands for OPortunistcally Elastic Resource Allocation. We have selected HTCondor for its maturity as a large-scale distributed system, aside, many other advantages. Thus, many European synchrotron and accelerator centres had recently moved toward HTCondor, as their cluster manager (e,g. CERN and DESY). Last

month I've been on a research visit to DESY at Hamburg, Germany and I had that idea of making OPERA more general platform than a Big Data bag of tasks only. Currently, the on going effort behind OPERA as a Container-based Cluster (CBC) can be described. There you can add any type of tasks in a container and run it at any type of resource.

SESAME take advantage of that in the following way. OPERA can run in three models. First, opportunistically or volunteer (using the idle and lightly used machines), which can be useful as a test pad for non-productive experiment and codes. Second, dedicated nodes (servers), which eliminate the “it works on the cluster” problem once and for all, by encapsulating the payload and its dependencies in a sandbox environment, without the need to install all applications, libraries, etc. at every single cluster node. Finally, OPERA-Provisioner that represent a hybrid platform of dedicated and non-dedicated nodes, aiming at enhancing the system throughput with minimal cost of deployment.

One can see many scenarios coming from collaborating to run high-throughput applications (LS Area E, image processing, and biological applications) on HPC infrastructure, using the containerization technology. What I am proposing is exploring the benefits of processing SESAME experiment tasks using OPERA and HTCondor frameworks.

OPERA-P aims to achieve improved performance and offer a reliable service on the proposed hybrid resource environment. While keeping in mind compatibility to the prior architecture. Apart from that, it is easy to see that the proposed POD approach is adaptable to many resource managers, such as Apache Mesos, Hadoop Yarn, and Docker Swarm, etc. Additionally, it is extensible to different BD framework, e.g., MapReduce (alongside with high-level interfaces Pig and Hive, etc.), Spark, Storm, Flink, etc. We also see that OPERA-P enables the system to scale out workloads, during high season periods. Without the need to add new servers or rent external cloud services, which optimize organizations idle resources utilization, and keeps cluster nodes busier due to the fine-grained nature of extending the tasks.

The beauty of this consolidation approach is not only allowing the system to run different workloads on an opportunistic cluster elastically but, make the system more dynamic and elastic than the prior design. This provisioning on-demand approach also makes it easier to expand resources given to the scheduler (whenever it is needed), without ever having to reconfigure the whole cluster. Hence, reduces operational costs, by utilizing these additional nodes, and the throughput will significantly improve. So, with a reduction of the total cost of ownership, and optimize idle resources utilization, this would imply as another direct impact and a greater return on infrastructure investments.

6.2.2.7 DICOM Network

DICOM Network
Life Sciences – Area E

Modern medical information systems integrate various types of medical equipment. This project should solve the actual problems for optimizing the processing and storage of medical

radiography investigations. The standard for working with medical images is the DICOM format, which allows storing studies in good quality with the patient's personal data included. The main problem in storing data in DICOM format are caused due to the fact that one study can take more than 1 gigabyte and consist of thousands of images.

The "DICOM Network" project goal is to provide access to investigations for medical staff with the appropriate access rights and as well as patients to the personal radiography investigations. At current stage eleven medical equipment’s are already connected to the system and now "DICOM Network" collects and processes more than 500 gigabytes of data per month. Flexible architecture of the system offers possibility to distribute Data Storage and Data Processing components between different processing units and storages, which could be customized using specific interfaces. The problem of storing medical investigations archive on national level can be considered as Big Data issue. Solution for this issue should take into account the different data access levels. On the one hand a medical investigation contains personal patient data, which means that data access should be restricted and secured. This could be reached by permission based categories of users and individual investigations access on supervised approval. On the other hand data should be accessible from any user, like patient or doctor, from any location. One of the main priority is system performance that should allow high speed access of the huge amount of data.

This system covers all necessary workflows for processing and documentation of medical investigations - from collecting images directly from equipment to archiving investigation in the patient medical record. “DICOM Network” offers extended functionality for enhancing quality of medical management and secured access to investigations. This helps doctors, specialists and penitents to gain access to structured database of medical images, allows documenting images that are collecting from various medical apparatus. At institutional level, the system helps to reduce costs of investigation, rise the quality of provided services.

"DICOM Network" is actively developing, but still far from realizing the potential built into the system. Taking into account the growing number of medical equipment and the trend towards modernization and computerization of health facilities, the system will be able to receive and will have to process dozens of terabytes of source information. The storage and subsequent transfer of such large amounts of data is an expensive process, impossible without optimization. On the other hand, for the successful development of the system it is necessary to provide the archive not for 3 years, as provided by law, but for tens of years to monitor the patient's condition and maintain a full medical record. It is also necessary to take into account the need for backup copies of such important information. It is easy to calculate that even for such a small country as Moldova, the data volumes are too large to store them in an unprocessed form. Thus, the issue of data optimization and archiving is a key factor for the development of such systems

6.2.3 Applications in Digital Cultural Heritage

6.2.3.1 DCH

The “Aharoni” Digitized Collections: Past, present and future of the southern Levant biodiversity
Digital Cultural Heritage – Area A

The “Aharoni” Online Digitized Collection is an innovative project aimed at creating a suitable platform for presenting and preserving the greatest Levantine faunal collection from the beginning of the 20th century. Natural History collections are rich repositories that document our planet’s past and present ecosystems and represent a monumental societal investment in research and applied environmental science. The National Natural History collections at the Hebrew University curate sole collections, The “Aharoni” collections, that are comprised of unique fauna (avian, amphibian, reptiles and mammalian) collections and archive material, are the sole direct evidence of the species richness and biodiversity of the Levant region at the beginning of the 20th century. The archival material includes historical documents describing Aharoni's numerous zoological expeditions through the Levant, his comments on specimens, ecological description of habitat that are gone, buried under cities and roads, and other valuable information associated with his studies. The southern Levant, a continental corridor between Europe, Africa and Asia, is a biodiversity hotspot and ideal natural laboratory for measuring historical evolutionary changes influenced by human civilizations in a faunal assemblage over time. Natural history collections represent a crucial contribution to the study of taxonomy, systematics, invasive species, biological conservation, land management and biotic responses to climate changes and human rapid alteration of the ecosystems in the region during the last 300 years. Unfortunately, these collections are significantly under-utilized due to difficulties in access to the specimens by the scientific community. Online digitization including 360-degree images of the specimens as well as the historical and current information on the species will serve as a high-quality database of the southern Levant fauna both for the academic community as a key biological resource and for the general public as a repository of knowledge on this unique region.

Parallel to the digitalization of the collection we would like to analyze the biological and historical data collected using the available analytic platforms in biodiversity science such as the Open Tree of Life, iDigBio, Lifemapper, Arbor and other complex post-tree analyses (e.g. niche modeling, niche diversification, and other ecological analyses). With the emerging cyberinfrastructure which provides unparalleled opportunities for mobilizing and integrating massive amounts of biological data, we believe that we can study the complex patterns conniving biodiversity loss in order to promote future land management and wildlife conservation programs.

6.2.3.2 mGeoAI

Massive Georeferencing of Aerial Images
Digital Cultural Heritage – Area C

Photogrammetry is quite an old discipline and even its digital version has a long tradition of studies and applications. Recently, thanks to the advent of free/affordable and easy to use amateur or professional applications, a number of scholars from various disciplines started to adopt photogrammetry on large-scale artifacts in Cultural Heritage: vases and statues, historical buildings, arriving to document entire landscapes.

The specific topic of photogrammetric applications to landscape studies constitutes probably one of the most important and promising research branch, i.e. giving archaeologists the possibility to get a digital measurable model of the object of their studies. Recent developments in the Computer Vision community offer new automated procedures for both image orientation and 3D reconstruction purposes at different scales. Complex scenes and objects can now be surveyed and reconstructed, using a large set of images with very satisfactory results. The final results are often completely compatible with (if not better than) laser scanning documentation, usually more expensive and time consuming in the field.

Despite the availability of a number of easy to use software specifically optimized for end-user computers (Graphic Processing Units or GPU), most of the image-matching processing is still a time consuming operation, and the required time exponentially increases with the number of images (and their sizes) to be processed.

The advent of drones (also known as Unmanned Aerial Vehicles or Remotely Piloted Aerial Systems) for low altitude aerial photography makes it even easier to collect endless number of overlapping photographs for different purposes and final ground resolution of the 3D model. These large amounts of data need proper processing especially when the required output is a set of georeferenced images. A specific software for automated image georeferencing has been developed at the GeoSat ReSeArch Lab IMS/FORTH by Gianluca Cantoro, allowing large image-set to be automatically matched with a given ortho-photo and geo-referenced in a completely automated way, without human input and completely controllable output. The software, AutoGR-Toolkit (now at version 3.5 http://www.ims.forth.gr/AutoGR) is already freely available online and it is consistently used by Universities, Research Centres, Commercial companies and private users all over the world (at the moment, more than 2000 users are counted). The possibility to have such a code on a much more powerful computer (or grid of computers) is of great importance, since large datasets can be easily and quickly georeferenced without dedicated computer and expensive hardware requirements for the end user.

6.2.3.3 3DGEOEX

3D non-invasive geophysical "excavation"
Digital Cultural Heritage – Area B

Ground based archaeological prospection methods have shown an exponential advance in Greece during the last 30 years in order to meet the needs regarding the efficient management of the Greek cultural heritage. All these years a significant experience has been gradually and systematically built based on the efforts of Greek universities and research institutes as well as foreign expeditions. Especially the research in archaeological prospection the last decade exhibited a slight deviation from the traditional mapping techniques to more sophisticated imaging and tomographic approaches to address specific problems.

Electrical Resistivity Tomography (ERT) comprises one of the most important modern imaging techniques since the technological advancements in hardware construction have the given the ability to acquire a large volume of data in limited time periods and at the same time the development of sophisticated inversion software resulted in the reconstruction a subsurface resistivity distribution which is consistent with the experimental data.

Despite the relatively routine employment of the ERT for reconstructing two-dimensional vertical sections of the subsurface, data processing and inversion can be extremely time consuming in cases of truly three-dimensional (3-D) approaches. Such cases involve millions of numerical calculations that also require extensive computer memory resources.

This project will focus on re-processing ERT data from past geophysical explorations in Greece. The access and use of the modern and updated inversion tools provided by VI-SEEM will contribute in revealing finer archaeological details hidden in the original data through the reconstruction of truly 3-D resistivity models of the hidden archaeological relics. The translation of the geophysical information to a virtual "excavation" will contribute in designing policies for the promotion and management of the archaeological sites.

6.2.3.4 VirMuF

Virtual Museum Framework
Digital Cultural Heritage – Area C

The Virtual Museum Framework (VirMuf) project is a set tools designed to work under the Unity game engine to help cultural heritage community create virtual museums in a fast and easy way. With cultural heritage digitization technologies becoming more robust and affordable, many museums worldwide are actively digitizing their collections. VirMuF is aimed towards how these digitized collections can be presented in an easy, meaningful and useful way for both professional and casual users. VirMuF makes it easy to create entire virtual museums with so many useful tools without having to do any programming. VirMuf is open-

source, and hence, users with programming knowledge can further extend it so as to fit their requirements.

6.2.3.5 DataCrowds

Data-Driven Crowds
Digital Cultural Heritage – Area A

Today more people are living in urban environments than in rural areas. It is forecasted that 70% of the global population will be living in cities by 2050. This intense urbanisation poses huge challenges in overcrowding, segregation, demographics and use of resources. The main goal of this project is to innovate in the unified area of research that is occupied with the transdisciplinary study of crowds in built environment. We envision a web-accessible, social platform that allows researchers from very diverse fields, such as Crowd Simulation, Urban Modelling and Simulation, Pedestrian Dynamics, Computer Graphics, Social Dynamics and Architecture to collaborate, share data and take advantage of each field's breakthroughs in order to contribute more accurate crowd simulations for the future sustainability of urban environments. As a first step in the implementation of this project, we will share tracked data of crowds from various sources in the Clowder platform of the VI-SEEM project.

6.2.3.6 HaPPEn

High Performing Photogrammetry
Digital Cultural Heritage – Area C

In the domain of reality based modelling, for almost 20 years, active sensors, such as laser scanners, have been the main devices available for creating accurate and detailed 3D models. They have been growing in popularity as a fundamental source of dense point clouds for 3D documentation, mapping and visualization purposes at various scales. Over the same period, until quite recently, photogrammetry and Structure from Motion methodologies were not able to efficiently deliver dense and detailed 3D point clouds similar to those produced by ranging instruments. MORE INFO FROM PI: A photogrammetric project, comprising a dataset of 1000 high resolution images and processed with intermediate resolution parameters, is completed within an average of 24 wall-clock hours on a WS intel(r) XEON(r) CPU e52680 v3 @ 2.50ghz 12 cores 24 logical processor 32 gb RAM. However with this WS specs it is anticipated that the photogrammetric task, processed at the highest resolution, may slow down (swap), increasing drastically the wall-clock time, or eventually fail due to lack of RAM memory. Core-hours time has to be assessed within the project and according to the minimum HPC hardware available for a proper and reliable comparison. All dataset, chosen for the benchmarks creation, will be processed both on the mentioned WS and the HPC architecture.

Thanks to recent significant improvements in hardware and algorithms, photogrammetry has reemerged as a competitive technology and a resurgence in automated photogrammetric methods is now evident. Image-based surveying and 3D modeling can now deliver results of comparable geometric characteristics to those of laser scanning for many terrestrial and aerial

applications. Therefore, the market, which was previously primarily dominated by airborne and terrestrial range sensors, nowadays offers more imagebased measurement tools for 3D recording and modeling (Remondino, 2014).

However due to the increasing dimension of massive imagery dataset, acquired from underwater, terrestrial and aerial platforms, the need of high performing computational resources has become fundamental in the scientific community, especially in the cultural heritage domain, where budget constraints and requests for high accurate models are addressed constantly. Commercial workstation and standard PCs indeed may not cope with the workload needed in terms of hardware and post processing time. Usually the final results represent a compromise between resources available and quality of the output. The implementation of photogrammetric tools on powerful computing infrastructures, which can be accessible remotely, freeing the user from any installation and configuration issue, is hence particularly timely in the research community, especially in the Mediterranean area.

The project will test and implement a set of commercial and open source software to be used for image based 3D reconstruction. Assessment on performance and usability will be realized, and benchmarks will be provided for further exploitation.

The tasks which will be analyzed are namely:

1 Features extraction
2 Bundle Block Adjustment (BBA)
3 Dense Stereo matching

To analyze the data, the results will be presented in terms of how long it takes each action to complete with X number of cores compared to how long it takes to complete with just a single core. From these results, the Amdahl's Law will be used to estimate the parallel efficiency for that action. 100% is perfect efficiency where a high core count CPU is ideal, but as the efficiency drops lower and lower having a high frequency CPU becomes more and more important. Tests are performed also on the GPUs performance assessing how each software benefits from their in comparison with CPUs computation only.

6.3 Application support

6.3.1 Service enablers

6.3.1.1 Responsibilities

Service enablers are responsible to implement the smooth integration of services into the VI-SEEM Virtual Research Platform (VRE). Each partner will assign a service enabler who will be responsible to coordinate and assist researchers from their own county during the process of the service integration. In particular each service enabler is responsible for the following:

1. Explain the integration procedure to researchers and ensure that the integration follows the agreed timelines as set in the integration plan.
2. Assist researcher's access to the VI-SEEM infrastructure and ensure the smooth initiation of service integration.
3. Provide technical support and address any problems encountered during the integration possibly with the support of other experts from the project. The service enabler together with the SC leader and WP5 leader will assign more experts (service integration team) to the project in case the service enabler does not have the full capacity to assist in all aspects of the project.
4. Ensure that researchers receive all required support from the partners that provide the computing resources.

All service enablers will be responsible to inform the WP5 and Task 5.4 leader about the progress of the integration and initiation of the services, and promptly report any problems. All the current service enablers are listed in Table 7.

Country	Name	Contact Information
Greece	Kyriakos Gkinis	kyrginis@admin.grnet.gr
Cyprus	Andreas Athenodorou	a.athenodorou@cyi.ac.cy
Bulgaria	Mariya Durchova	mabs@parallel.bas.bg
Serbia	Dusan Vudragovic	dusan@ipb.ac.rs
Hungary	Tamas Maray	vi-seem-support@niif.hu
Romania	Silviu Panica	silviu.panica@e-uvt.ro
Albania	Neki Frasheri	nfrasheri@fti.edu.al
Bosnia and Herzegovina	Mihajlo Savic	badaboom@etfbl.net
FYR of Macedonia	Anastas Mishev	anastas.mishev@finki.ukim.mk
Montenegro	Luka Filipovic	lukaf@ac.me
Moldova	Alexandr Golubev	galex@renam.md
Armenia	Wahi Narsisian	wahi@sci.am
Georgia	Temur Maisuradze	temur@grena.ge
Egypt	Youssef Eldakar	Youssef.Eldakar@bibalx.org
Israel	Zivan Yoash	zivan@iucc.ac.il
Jordan	Salman Matalgah	salman.matalgah@sesame.org.jo

Table 7 – The list of Service Enablers

It is clear that the service enablers play an important role during the integration phases. Nevertheless, service enablers also contribute importantly towards the open calls. In the context of the 1st and 2nd open calls each service enabler is responsible for the following:

1. The service enabler contacts the primary investigator and guides him/her through the process of getting access to the VI-SEEM resources which have been allocated through the review process.
2. For any technical questions regarding the initiation of the project the primary investigator refer to your service enabler.
3. After the primary investigators gain access on the allocated VI-SEEM resources and services, whenever they experience a trouble or issue they first contact the service enabler to address the trouble.

The 1st and 2nd calls for proposals for projects accessing VI-SEEM resources and services and the service enabler responsibilities at all stages of the call are outlined in Figure 1.

6.3.2 Software

Applications that requested computational time either in HPCs or Cloud or Grid, require the existence of particular software tools which must be installed in the machines. Hence, the allocation of resources took into account the software requirements for each application and assigned the project according to the software needs to the right machines. In case that a machine requires the installation of a software, this is done with the support of the local contact site of the machine. Lists with the software required for each application is provided in Table 8 and Table 9 for the 1st and 2nd Open Calls respectively.

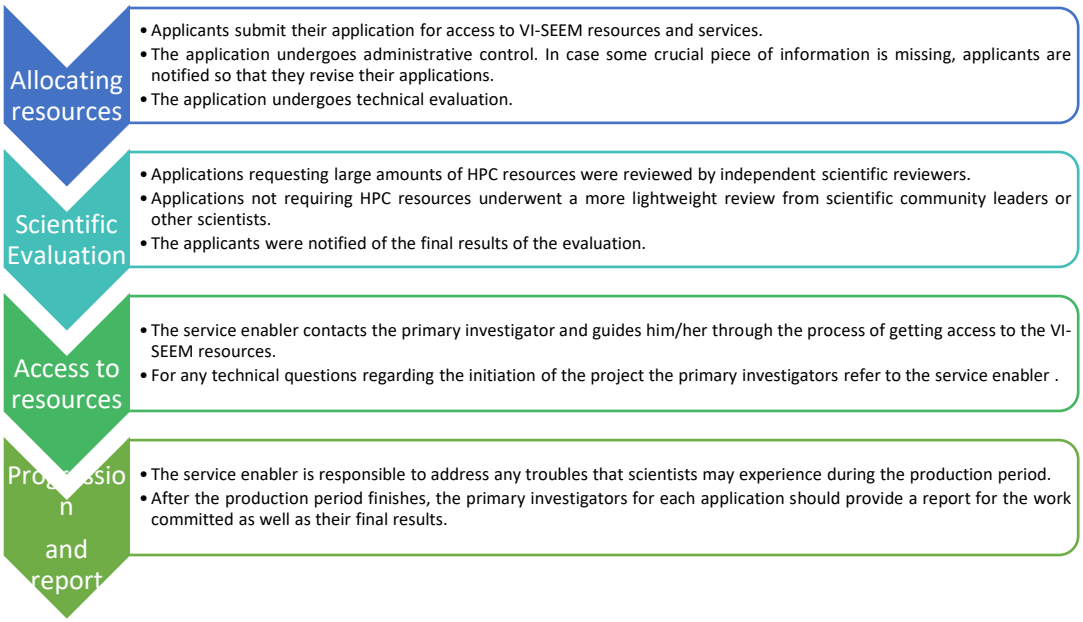


Figure 1 - The flowchart for the integration of a service in the VI-SEEM VRE

Application Acronym	Required Software
VINE	WRF-Chem, RegCM4
PSOMI	N/A
MECCA_ECHAM_GPU	MESSY/MECCA/ECHAM
TVRegCM	RegCM
ACIQLife	WRF, CMAQ
CH-CBIR	CH-CBIR
Continuous_LST	N/A
MULTIDRUG	Gromacs, NAMD
DATANMNH	N/A

D3R	Schrodinger FEP+/Desmond GPU
MC4CH	N/A
MolSurf	Gromacs, VMD
DRS-ACS	Gaussian09, CPMD
CNCADD	Gaussian09, CPMD
BVL	N/A
DREAMCLIMATE	DREAM
AUTH_WRF_371_M	N/A
DACEM1	WRF-Chem
WRF-ARW	WRF
DACEM2	WRF-Chem

Table 8 – The required software for each selected application of the 1st Open Call

Application Acronym	Required Software
DCH	CultLab3D
mGeoAI	AutoGR-Toolkit, OpenCV3, PIL, pyproj
3DGEOEX	3DINV
BRING-MD	VMD, NAMD,
OP4D	Generic 3D mesh processing
KPP_CLIMATE_GPU	MESSY, MECCA, ECHAM
WheAirCYEM	WRF-Chem
SIPD	GROMACS
VirMuF	Self-developed code
3DVAR_WRF	WRF
Surf_prop	GROMACS, VMD, GAMESS-US
CCCSCMRR	Self-developed code
OPERA-P	IMAN1-Booster/King
NBBM4RHMS	NEMS-NMMB
DataCrowds	-
DICOM Network	Self-developed code
HaPPEn	MicMac, Agisoft Photoscan

Table 9 - The required software for each selected application of the 2nd Open Call

7 Results of the Applications

7.1 Applications from the 1st Open Call

7.1.1 Applications in Climate research

7.1.1.1 VINE

VIAM/NEA Regional Chemistry-Climate Model
Climate research – Area A

In the frame of VI-SEEM project WRF-Chem model was compiled and installed on the GRENA cluster (scripts for automation of data download, pre-processing and run the model were created). Also there were preliminary installed required software libraries (HDF5, NetCDF, jasper, png and zlib). Researchers have used WPS and WRF-Chem v3.7.1 with basic nesting option. For the Compilation of the model Intel Fortran/C compilers were chosen. With the purpose of preliminary examination, the WRF-Chem model during the Chemistry run used 1 nested configuration domain. For the purpose of obtaining initial data for WRF-Chem model the AI data from TOMS and OMI instruments were available from the NASA Goddard Earth Sciences Data Information Services Center (GES DISC) (http://acdisc.gsfc.nasa.gov/). Above the territory of Georgia, the ground observations are available from 7 stations, where dust particles in air are measured 3 times a day, and its comparison with satellite data have been made.

On the bases of WRF v.3.6 model configured for the Caucasus region, two particulate cases of unexpected heavy showers which took place from 2015 to 2017 in Tbilisi and in Kakheti (eastern Georgia) have been studied. Simulations were performed by two set of domains with horizontal grid-point resolutions of 6.6 km and 2.2 km. The ability of the WRF model in prediction precipitations with different microphysics and convective scheme components taking into consideration complex terrain of the Georgian territory has been tested. Some results of the numerical calculations performed by WRF model are presented in the published articles.

The Regional Climate Model RegCM v4.3 was used for climate simulation over Caucasus domain. Model was applied for simulation of 1990-2002 period with ERA Interim reanalyses for model tuning and optimal configuration. Model was configured for two land model (bats and LCM) options with different combination of skims and nesting opportunity with horizontal grid mesh 10 km have been examined.

Study was focused on model fidelity of precipitation performance. Model output (SRF data) was compared to E-OBS gridded data, as well as for nested domain point based (observed time series) have been used. Fulfilment of such tasks requires High Performance Computing and quite big storage space, various programs and scripts for manipulating on different type formats and projections.

Essential Part of this study has been performed on GRENA cluster with 2 nodes: each of them with 2 X Intel Xeon E5-2670 8C 2.6 Ghz. 32 GB RAM, 10 Gb/s Ethernet and 8TB storage. Most of post processing scripts and programs have been created and tested there. For long period climate simulations, the Supercomputer System Avitohol at the IICT–BAS provided a great opportunity. Which is built with HPC Cluster Platform (426 nodes). Each node consists of 2 X (Intel Xeon E5-2680v2 10Cores 2.8GHz CPUs) (total 852 CPUs), non-blocking InfiniBand FDR, AND 18 phi nodes, each of them with 2 x INTEL Xeon Phi 7120p and the storage is provided by a storage system with 250 + 500 TB of raw disk storage capacities. The workflow and some of the produced outputs were uploaded to be used by the community and the other partners of the project VI-SEEM on the repositories site which is hold by GRNET. All applications: WRF; WRF-Chem; RegCM4 models and visualization tools – GRADS and NCL have been installed on the GRENA and ARIS clusters based on EMAC’s boundary conditions. The differences in wet days number, also for precipitation sums are maximal in summer. Model overestimates both parameters for all seasons and over the whole domain, but for other seasons it in the range of 6-9 days. Each picture has orographic features, where differences are dominant and model tend to increase local convective processes.

For further improvement of model performance better investigation of sub-grid scale processes, more cross validation against more sources of observation such as satellite will be benefiting. Bias and noise reduction in model results can make it reliable for future prediction. Besides VINE team was focused to evaluate regional climate model output. For this purpose, several additional tools and environments have been used. First of all researchers have installed the latest version of RCMES Virtual Machine (VM) – an emulation of a computer system on the desktop PCs. RCMES is packaged as a Virtual Machine image that contains all of the software, dependencies, libraries, configuration files, sample observation and model datasets, etc. and was needed to quickly get up and running with a model evaluation experiment. The virtual machine packaging dramatically reduces the time required to obtain the necessary environment for experimenting with RCMES. RCMES supports a wide variety of use cases, from simple non-connected tasks like performing re-gridding of model and observational data, to a complete model evaluation workflow from configuration to output plots. RCMES provides convenient access to multiple sources of data:

- Gridded satellite observations of key physical variables curated for query and subset from the RCMED database;
- Download of model runs or gridded observations (obs4MIPS) stored in the Earth System Grid Federation (ESGF);
- Local netCDF files containing model variables or satellite observations provided by the investigator.

The OCW readers can ingest data on regular grids (lat/lon/alt coordinates are vectors) or high-resolution irregular grids (lat/lon coordinates are a 2-dimensional mesh). For netCDF files that satisfy the regional the Climate Forecast (CF) metadata conventions, attributes and metadata are automatically recognized and extracted (offset, scale, units, etc.). Adding another reader for a specific file format can usually be easily done in Python.

Using RCMES package the researchers performed compression and evaluation of existing RCM runs for Caucasus territory for near surface mean temperature and precipitation and started perform clouds evaluation. Besides RCMES tool and packages such as CDO, NCL and NCO was used. Some results of those finding are described in article “Effect of dust aerosols on in forming climate of Georgia”.

One of the main scientific importance of the project’ s expected outcomes is the creation of the state-of-the-art multi-model regional climate prediction ensemble system, with modern computational and communication technologies for the Caucasus region. Expected outcomes of the investigations in the prediction of regional climate change and its impacts on socio-economic sectors at different timescales periods will have a great scientific importance not only for scientific, environmental, economic, agriculture, industry and tourist communities on the territory of Georgian but for agencies and companies of the neighbouring countries of the Caucasian region. Outcomes of these investigations in the studding of uncertainties in emissions and climate parameters (parameterization of different climate parameters, and reduction of uncertainties in RCMES) have a great scientific importance in the field of regional climate prediction by numerical modelling and as well as in studying the climate change field. Outcomes of these investigations in understanding and clarifying the cause and effect of the increased frequency of severe droughts, heavy showers and growth of the Caucasus glaciers melting on the territory of Georgia have a great scientific importance for the scientists, analysts of the administration and private business (industry) makers on the Caucasus territory. Also as these outcomes will be based on the high resolution, fully comprehensive, ensemble modelling system so expected outcomes of our investigations, undoubtedly will be useful for scientists working in the field of regional and global climate change.

Figure 2 from provides an example of the Precipitation bias in RegCM model simulations for South Caucasus Region while Figure 3 demonstrates biases in mean seasonal number of wet days and mean seasonal precipitation sums for the mean of different model configurations.

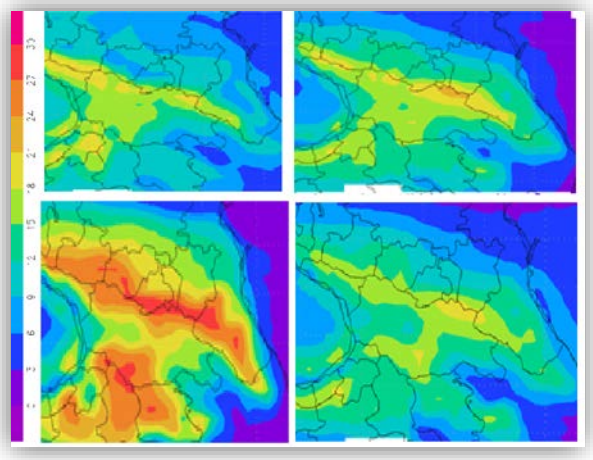


Figure 2 – Precipitation bias in RegCM model simulations for South Caucasus Domain

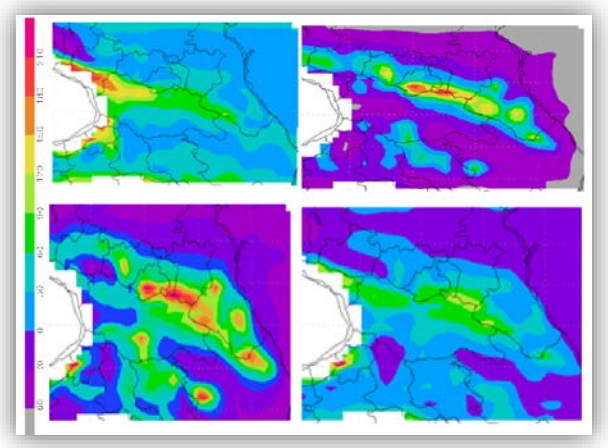


Figure 3 – Biases in mean seasonal number of wet days (upper) and mean seasonal precipitation sums (bottom) (units: mm) for the mean of different model configurations.

7.1.1.2 MECCA

MECCA / ECHAM GPU Acceleration
Climate research – Area A

The outcome of this study was the development of a publicly released source code that automatically ports the kinetics calculations on GPU architectures. Each GPU thread calculates the chemical concentrations of an individual atmospheric grid box, massively parallelizing the workload.

The researchers developed a software package that automatically generates CUDA kernels to numerically integrate atmospheric chemical kinetics in the global climate model ECHAM/MESSy Atmospheric Chemistry (EMAC), used to study climate change and air quality scenarios. A source-to-source compiler outputs a CUDA-compatible kernel by parsing the FORTRAN code generated by the Kinetic PreProcessor (KPP) general analysis tool. All Rosenbrock methods that are available in the KPP numerical library are supported.

Performance evaluation, using Fermi and Pascal CUDA-enabled GPU accelerators, shows achieved speed-ups of 4.5 × and 20.4 ×, respectively, of the kernel execution time. A node-to-node real-world production performance comparison shows a 1.75 × speed-up over the non-accelerated application using the KPP three-stage Rosenbrock solver. The researchers provided a detailed description of the code optimizations used to improve the performance including memory optimizations, control code simplification, and reduction of idle time. The accuracy and correctness of the accelerated implementation were evaluated by comparing to the CPU-only code of the application. The median relative difference is found to be less than 0.000000001 % when comparing the output of the accelerated kernel the CPU-only code.

The approach followed, including the computational workload division, and the developed GPU solver code can potentially be used as the basis for hardware acceleration of numerous geoscientific models that rely on KPP for atmospheric chemical kinetics applications. A schematic representation of the GPU Tasks offload execution of atmospheric chemical kinetics calculations is provided in Figure 4.

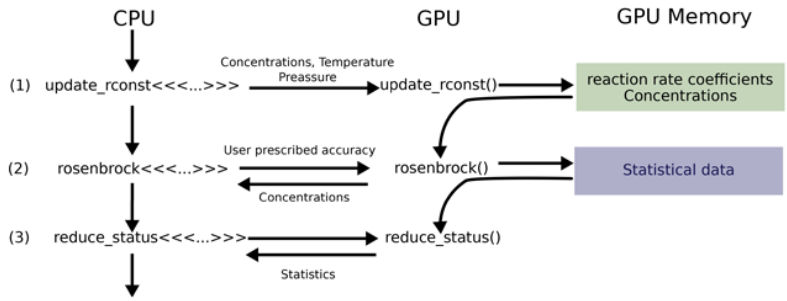


Figure 4 - GPU Tasks offload execution of atmospheric chemical kinetics calculations.

7.1.1.3 TVRegCM

Tuning and Validation of the RegCM
Climate research – Area B

The application TV RegCM allowed the research team to run computer simulations for a period of 10 years at 10km grid resolution and for 20 different model configurations as well as to choose the best model performance, according to the surface temperature and precipitation. It became clear that not all of parametrizations could run at this resolution (they became unstable) and some were not included into these 20 different model configurations. Generally it could be concluded that the model configurations with Grell cumulus parametrization schemes are performing better than the others.

This investigation wouldn't be possible be completed without using the computational resources of VI-SEEM. An extensive database was created from the computer simulations and the produced results can be used as basis in the study of future climate change and its impact on ecosystems, economics and the quality of life.

Examples of results obtained within the scope of TVRegCM can be seen in Figure 5 and Figure 6.

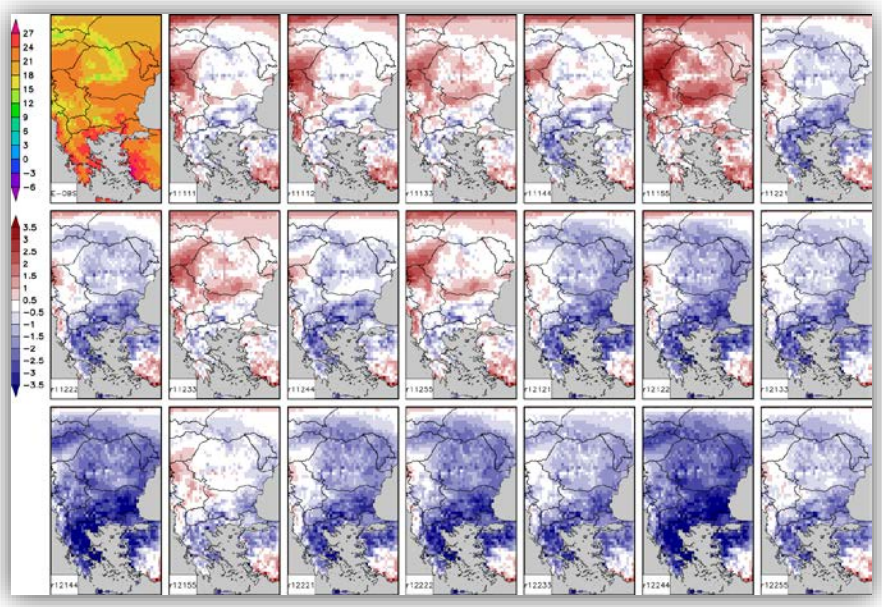


Figure 5 – Mean summer temperature (upper leftmost subplot, units: °C) and biases (units: °C) of the considered model configurations

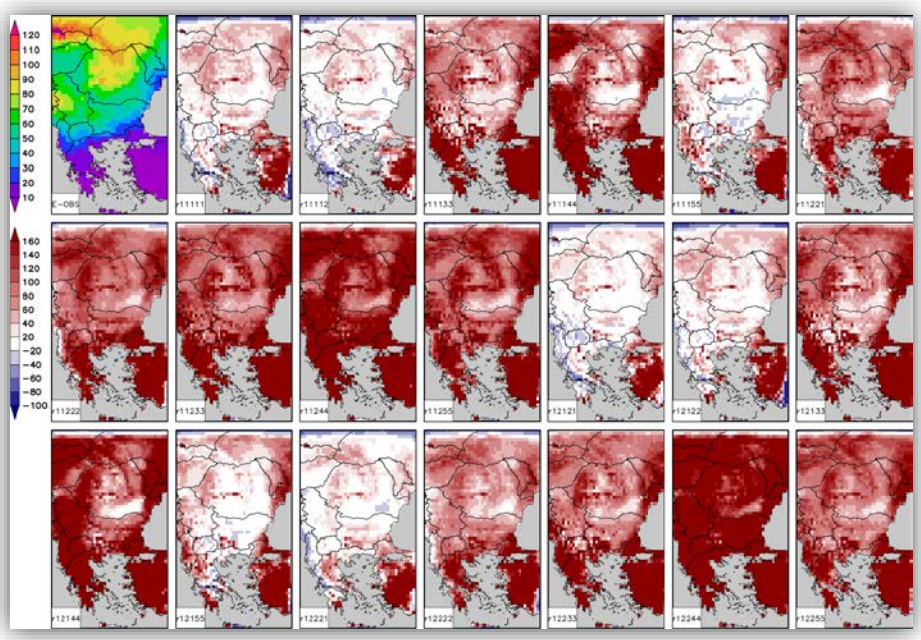


Figure 6 – Mean summer precipitation sum (upper leftmost subplot, units: mm/month) and relative (in %) biases of the considered model configurations

7.1.1.4 ACIQLife

Atmospheric Composition Impact on Quality of Life and Human Health
Climate research – Area B

Calculation of the AQ impact on human health and quality of life in Sofia city is the objective of this study. The impact is calculated in the terms of the so called Air Quality Indices (AQI) – an integral characteristic directly measuring the effects of AQ on human health. The calculations are made on the basis of long term AQ simulations, which make it possible to revile the climate of AQI spatial/temporal distribution and behaviour. The AQI is defined as a measure of air pollution seen in the context of its impact on human health. It provides an integrated assessment of the impact of the whole range of pollutants on human health and is calculated based on the concentration of various pollutants obtained from measurements or numerical modelling. The simulations aimed at constructing of ensemble, comprehensive enough as to provide statistically reliable assessment of the atmospheric composition climate in Bulgaria and Sofia – typical and extreme features of the special/temporal behaviour, annual means, seasonal variations, etc. The computer simulated ensemble is utilized for studying calculation of AQI, constructing the AQI climate and analysis of its spatial - temporal behaviour. The results from the simulations shows the AQ status of the areas of interest, shows dominant pollutants and probably reasons and sources of pollutants. Could be follow contributions of different dynamical and chemical processes and different sources to the surface concentrations of pollutants, respectively to the AQ status and health status for the general population. The obtained results can be used in the formulation of short-term (ongoing) solutions and long-term strategies to reduce air pollution.

Very extensive database was created from the numerical simulations, which could be used for different studies and considerations of the main features and origins of the atmospheric composition in different scales, including the AQ climate.

Thanks to the opportunity of HPC platforms was possible for detailed and extensive study of the atmospheric composition – its behaviour, origin and health impact. The generated ensembles of atmospheric composition characteristics have still to be carefully and extensively treated and analysed, which will be objective of the future work.

Selected results are presented below. Namely, Figure 7 represents the annual plots of the recurrence while Figure 8 demonstrates the annual diurnal variations

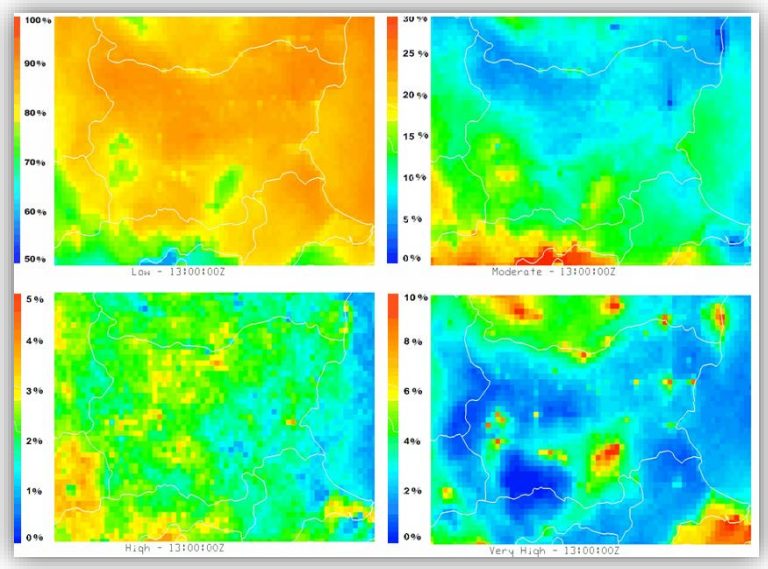


Figure 7 – Annual plots of the recurrence [%] of the AQI - Low, Moderate and High bands in Bulgaria.

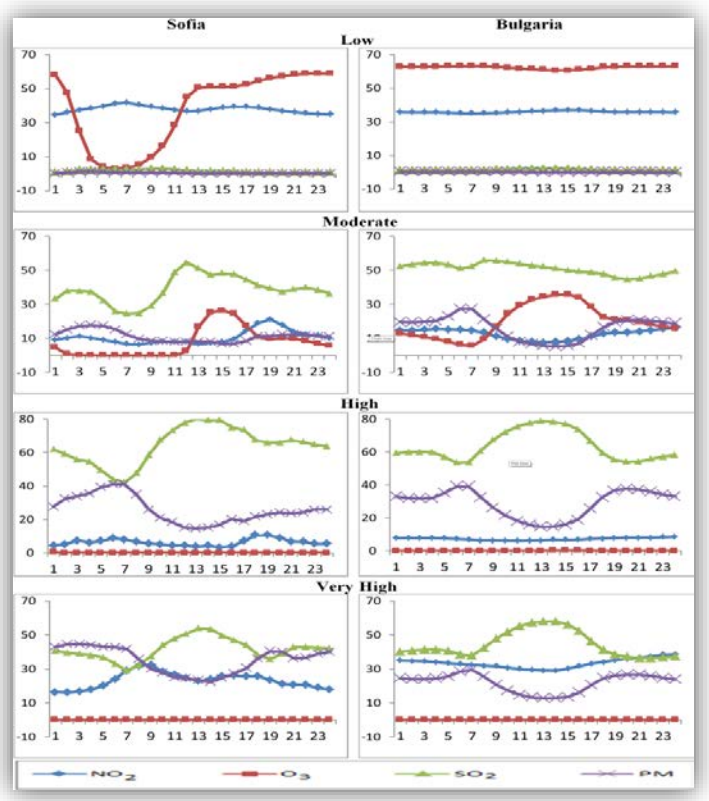


Figure 8 – Annual diurnal variations (%) of the dominant pollutant.

7.1.1.5 Continuous_LST

Continuous Land Surface Temperature
Climate research – Area A

Model performance: Figure 9 shows the Continuous LST for the year 2010 together with the original MODIS LST observations time series, showing a good agreement between the two (r2=0.90, p<0.0001, n=245 days). Good agreement between continuous LST and MODIS LST is also displayed for all cloud free pixels in the entire study area for 2003-2010 (r2 = 0.89; Figure 10a). For comparison, the correlation between MODIS LST and Tclim results in a lower correlation (r2=0.87, not shown) and significantly higher bias (-0.18°K compared to -0.04°K) and RMSE (3.5°K compared to 2.6°K) demonstrating the important contribution of the anomaly component in the continuous LST. The anomalies from MODIS and CFSv2 show a similar pattern but with a wider temperature range from MODIS (-13 to 12°K) compared to that from the CFSv2 model (-7 to 10°K) (Figure 10b).

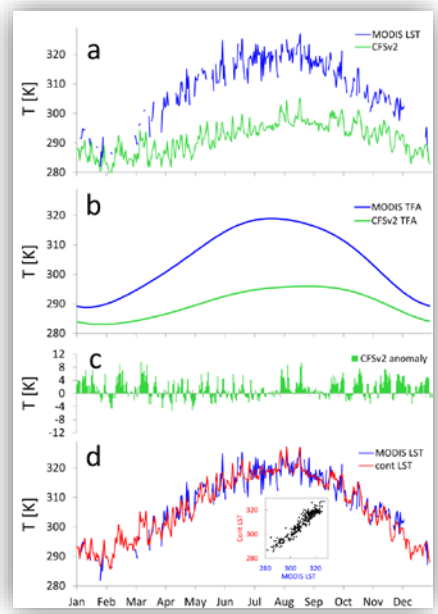


Figure 9 – Surface temperature of one year (2010) from MODIS, CFSv2, their climatological values (TFA), and the model (cont LST) from a single pixel (lat lon).

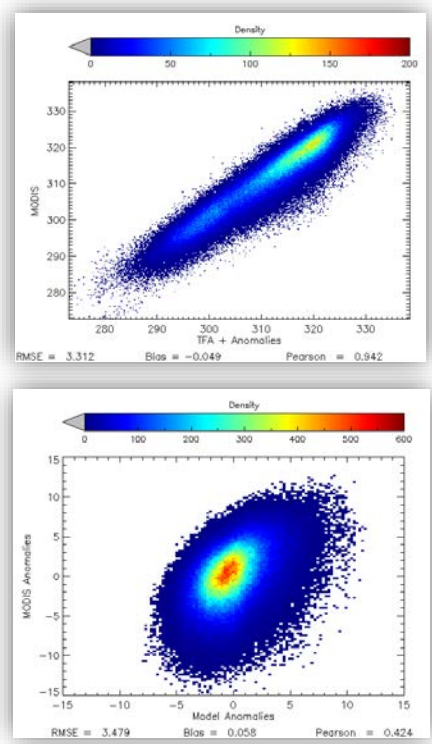


Figure 10 – Scatterplots of (a) the modeled vs. MODIS daily LST and (b) the CFSv2 (Fig 2c) vs. MODIS daily LST anomalies (actual minus climatological temperatures) for 2003 to 2010 in the study area.

The effect of synoptic circulations on continuous LST: The researchers demonstrate the cloudy pixel filling method on two contrasting synoptic configurations, high to the East and deep low to the East, which occur from mid-fall to mid-spring (Figure 11a, b). These synoptic circulations are associated with high cloud cover and significant missing LST data (Figure 11c); therefore, making it relevant to demonstrate the method (Figure 12 a, e).

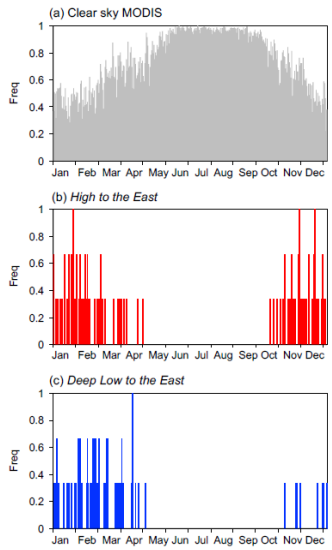


Figure 11 – Histogram for synoptic circulations on continuous LST

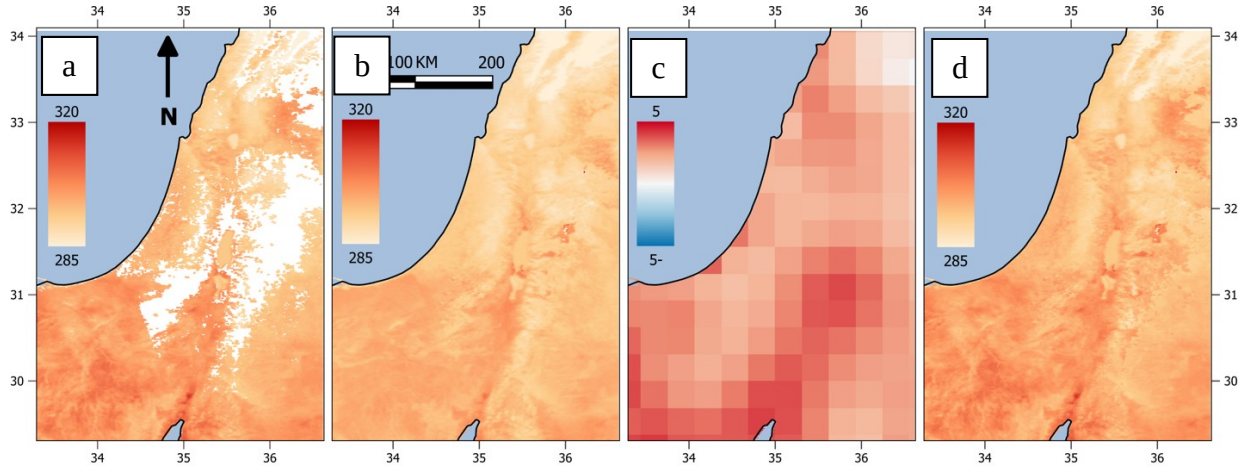
The High to the East synoptic category, is a surface high pressure developing over Syria and Iraq. This high pressure generates south to south-easterly winds that advects warm air towards the EM resulting in positive temperature anomalies (Figure 12c) with respect to the climatological LST for the specific day of year (Figure 12b). The result of the continuous LST is presented in Figure 12d showing warmer LST than the climatological mean (Figure 12b). The Deep low to the East synoptic category, is related to a low barometric trough centered over Iraq generating a north-westerly flow over the EM. This flow advects cool air from the north resulting in strong negative temperature anomalies (Figure 12g), resulting in lower continuous LST (Figure 12h).

During the summer, the dominant synoptic circulation prevailing over the EM is the Persian trough, which is associated with clear sky conditions and high frequency of cloud free MODIS pixels (Figure 11a), making it irrelevant to the continuous LST method.

A global continuous LST is a vital product variable in climatological and environmental studies. This project can be used for monitoring vegetation water stress, assessing surface energy balance, detecting land surface disturbance, and public health (exposure assessment and monitoring condition suitable for vector-borne disease) and climate and climate change related studies.

In future work, the Continuous LST product will be used in the field of cloud detection in satellite imagery.

High to the East



Deep low to the East

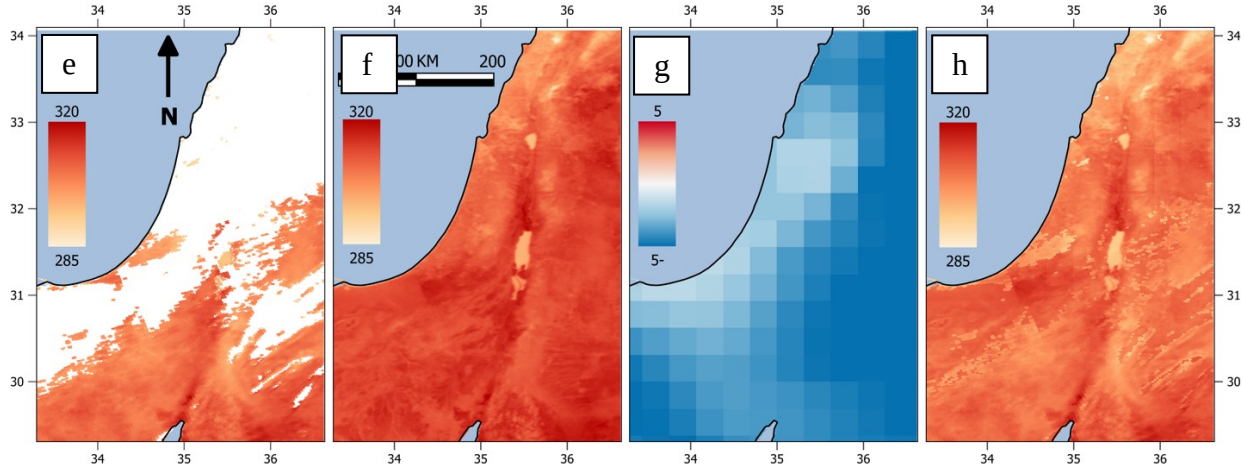


Figure 12 – Maps of (a,e) original MODIS LST (b,f), climatological LST (c,g), the temperature anomalies from CFSv2, and (d,h) cont LST in the EM for two typical High to the East (a-d) and Deep low to the East (e-h) events. These two specific events occurred in the winter (January).

7.1.1.6 DRS-ACS

Dynamics, reactivity and spectroscopy of atmospheric chemical species
Climate research – Area B

Researchers working on DRS-ACS have computed a number of properties relevant to detection and identification of atmospherically relevant individual molecular species and non-covalently bonded clusters that they form under realistic conditions encountered in the Earth's

atmosphere. Computations have been performed within the “static” and “dynamic” approaches, most of which were non-standard (i.e. non-routine), i.e. they went beyond the usually applied normal mode approximation (NMA) in computing vibrational spectroscopic properties of molecular species (at 0 K). The main advantage of the dynamical approach in this context is the possibility to assess the finite-temperature spectral properties of molecules relevant to climate science and atmospheric chemistry, beyond the harmonic approximation, inherent to NMA.

The first considered system was the simplest organic acid which is, at the same time, the most abundant organic molecule in the Earth’s atmosphere. Vibrational spectra of the two conformers of the free formic acid molecule were computed by various static and dynamic approaches, with a special emphasis on the region of O-H stretching modes. The static approaches were based on sequential computation of anharmonic O-H stretching vibrational potentials and numerical solution of the vibrational Schrödinger equation by the Numerov method or linear variational approach. The dynamic approach was based on molecular dynamics (MD) simulations performed within the atom-centered density matrix propagation scheme (ADMP) or within the Born-Oppenheimer MD (BOMD) scheme, followed by analysis of certain time-correlation functions, computed from the equilibrated ADMP/BOMD trajectories. In the beginning of the dynamical simulations, researchers have supplied each system with a predefined amount of nuclear kinetic energy, which was further redistributed among the modes; further propagation of the system’s dynamics was achieved either by the atom-centered density matrix propagation scheme (ADMP), or by achieving full convergence of the SCF equations in each step (i.e. by the BOMD scheme). In this way, they have followed the classical dynamics of nuclei on the ab initio, DFT or semi-empirical potential energy hypersurface (PES) generated “on-the-fly”. As all of the simulations have been carried out within the NVE ensemble, varying the system’s temperature (by the amount of initially supplied nuclear kinetic energy) it is allowed to sample more and more extended regions of the molecular PES. Finite-temperature spectra were computed by three approaches: i) by Fourier transformation of the velocity-velocity autocorrelation functions computed from the productive (equilibrated) MD trajectory; ii) by Fourier transformation of the dipole moment autocorrelation function; iii) by Fourier transformation of some of the most relevant partial correlation functions computed for particular local or normal modes. Researchers have computed various types of spectra at series of temperatures (ranging from 10 up to 300 K), including the rovibrational density of states spectra (derived from the velocity-velocity autocorrelation functions), as well as the infrared absorption cross section spectra (computed from the dipole moment autocorrelation functions). The thermally induced changes in the single-molecule spectroscopic properties were deduced and the reasons behind them were analyzed and discussed. The advantages of the dynamic approach to computational vibrational spectroscopy at finite temperatures were outlined and discussed, in the context of research devoted to molecular-level understanding of the phenomena and processes relevant to climate science and atmospheric chemistry. Calculations for this system have been carried out within the density functional tight binding (DFTB) formalism, as well as at the second-order Møller-Plesset perturbation theory level, using the 6-311++G(3df,3pd) basis set for

orbital expansion (MP2/6-311++G(3df,3pd)). The computed properties were compared to the available experimental data.
For comparison purposes, in parallel to the dynamical approach, the complete O-H stretching spectral bands were also computed by carrying out “static” computations of the 1D O-H vibrational potentials for series of “snapshots” taken from the equilibrated MD runs, chosen in a way that the mutual statistical correlation is minimized. They referred to this technique as “sequential MD + QM approach” (S-MD-QM). Certain differences between the results from the S-MD-QM and the dynamical approach were analyzed and discussed from the viewpoint of fundamental statistical mechanics principles.

Analogous computations have been carried out for the two possible conformers of 1-naphtol. Particular attention was paid to the peculiar situation of intramolecular confinement of the O-H stretching motion in the case of this molecule when the higher-energy cis conformer is in question. Aside from computation of the static (0 K) and temperature-dependent vibrational density of states spectra, extensive natural bond orbital analyses of the O-H stretching motions in the two conformers of 1-naphtol have been carried out as well. They have obtained excellent agreement between our theoretical prediction and experimentally measured vibrational frequency shift of the O-H mode upon cis-trans conformational interchange. Since the computed (nuclear) trajectories are essentially classical, the possibility for intramolecular torsional tunnelling has been accounted for in a-posteriori manner, employing the quasi-classical (WKB) approximation. Approaches for exact treatment of all these issues have been developed and implemented that, as mentioned before, enable theoretical studies under realistic conditions, often realized in the actual experiments. This also allows even better and more in-depth analysis of already published experimental results and often their reassignment or even complete reinterpretation.

Aside from the free formic acid molecule, also in-depth theoretical modeling of its complexes with benzene has been carried out, implementing the previously described static and dynamic approaches. Special emphasis was again put on the O-H stretching mode as a vibrational chromophore. The theoretical predictions were again used to further rationalize and even provide more in-depth insights into the available spectroscopic data. For example, the theoretical approach enabled to predict that both conformers of formic acid are capable of interacting noncovalently with the π-electronic cloud of benzene molecule (forming an O-H⋯π hydrogen bond), therefore allowing further possibilities for conformational isomerism in the formic acid … benzene dimers.

The research efforts were further directed towards in-depth theoretical studies of magnetic shielding properties of various water and hydroxide clusters of Al(III) ionic species by the sequential MD + QM approach, as well as by analogous Monte Carlo + QM methodology (MC + QM). Our theoretical predictions of these properties are in excellent agreement with the state-of-the-art experimental data (within the experimental error). Some pictorialisations of the results are presented in the following Figures 13- 16.

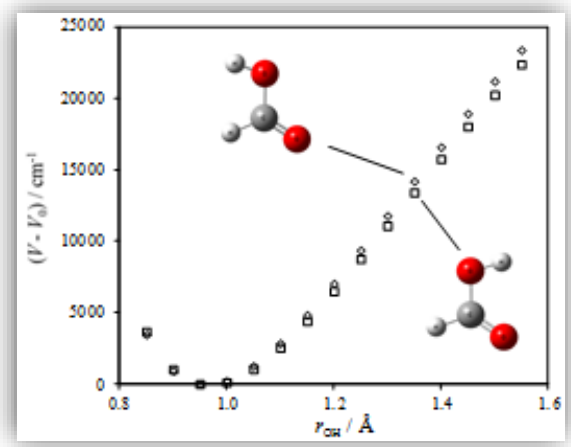


Figure 13 - Typical O-H stretching vibrational potentials computed at second-order Møller-Plesset perturbation theoretic level (MP2/6-311++G(3df,3pd)) for cis- and trans- conformers of formic acid. In the course of our approach, these can be computed either for the geometries corresponding to the minima on the same potential energy surface (for the static, temperature-independent case), or for series of geometries extracted from BOMD snapshots (this will give the realistic, temperature dependent theoretical spectral band, comparable to experiment).

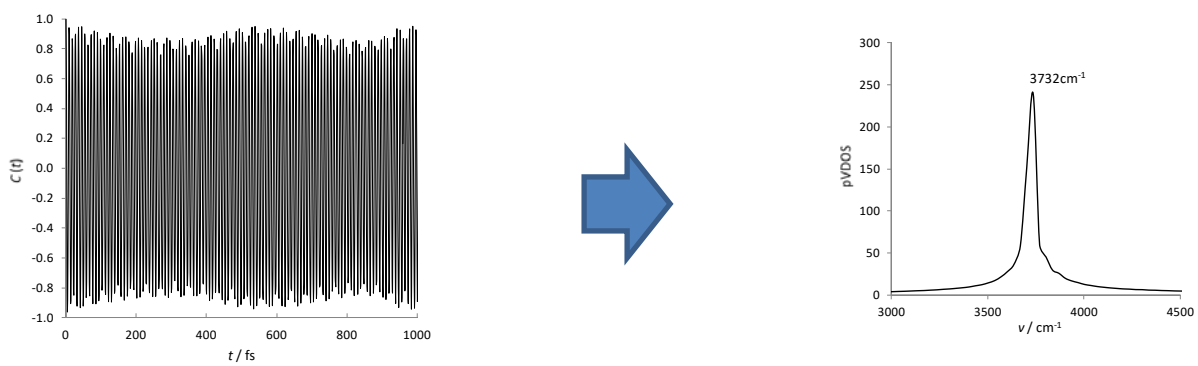


Figure 14 – The partial autocorrelation function (pACF) for the O-H stretching coordinate time derivative computed from BOMD/MP2/6-311++G(3df,3pd) trajectory at 10 K for cis- conformer of formic acid and the corresponding partial vibrational density of states (p VDOS) computed by sequential Fourier transformation of C(t).

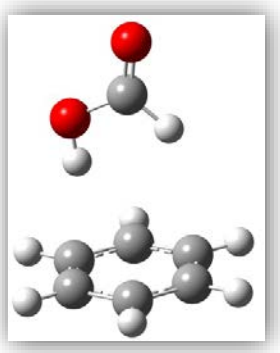


Figure 15 – The minimum on the B3LYP/6-311++G(3df,3pd) PES of the cis-formic acid – benzene dimer

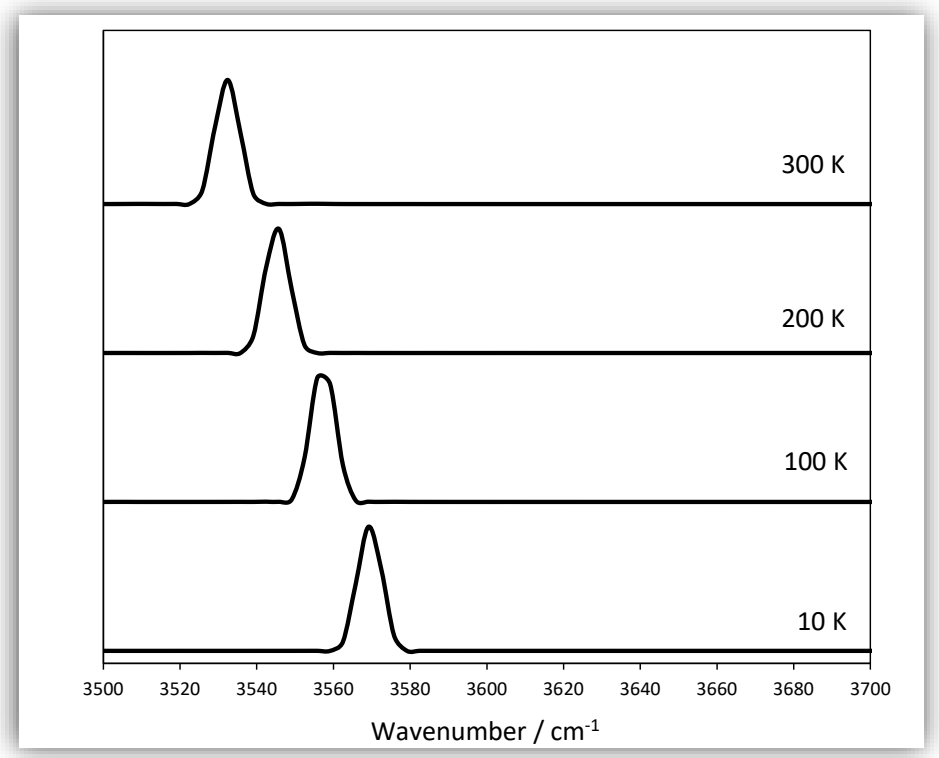


Figure 16 – The kinetic energy spectra (kinetic energy density of states spectra) obtained by Fourier transformation of the velocity-velocity autocorrelation function for the series of simulations carried out at the four different temperatures, starting from the trans-conformer of the free formic acid molecule, in the frequency region where signals due to the O-H stretching vibrations are expected to appear (from 3500 to 3700 cm^{-1}).

7.1.1.7 DREAMCLIMATE

Dust Regional Atmospheric Model Climatology
Climate research – Area B

Using the DREAMCLIMATE service at PARADOX during the first VI-SEEM call for production use of resources and services, the researchers produced a dataset with the aerosol optical thickness and surface dust concentration for the one-year period. They selected the year 2005 for this analysis, which serves as an example and demonstrates usability of DREAMCLIMATE service. The dataset cover the wide region of North Africa, Southern Europe and Middle East in 30 km horizontal resolution with 28 vertical levels, and is made publicly available via the VI-SEEM data repository.

In addition to this initial dataset, they also produced a dataset with a higher resolution of 15 km for the same region and period of time. The global mean DREAMCLIMATE-modelled dust concentration for year 2005 is presented in Figure 17.

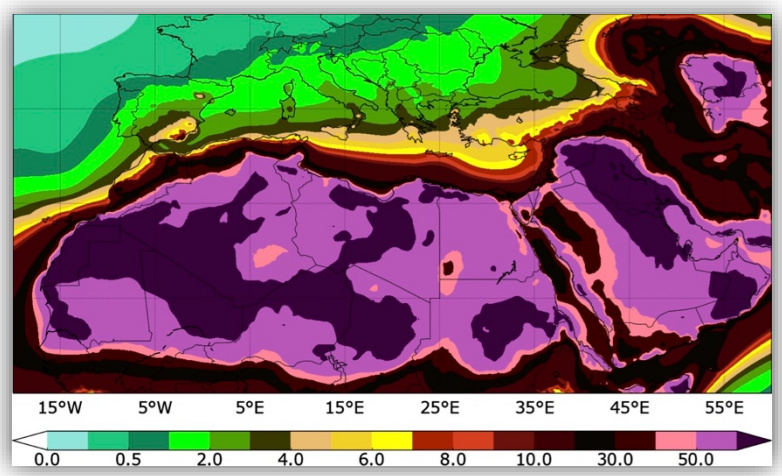


Figure 17 – Calculated mean dust concentrations in µg/m3, obtained from the DREAMCLIMATE model. The model integration area covers region of North Africa, Southern Europe and Middle East, with 15 km horizontal resolution in 28 vertical levels for the year 2005.

Using the human health impact function, the researchers related the changes in pollutant concentrations to the changes in human mortality, and estimated the global annual premature mortality due to airborne desert dust. For this, they use as a baseline the mortality rate estimated by the World Health Organization (WHO) Statistical Information System on the country-level based on the International Classification of Diseases 10th Revision (ICD-10) classification, and regional data from the WHO Global burden of disease for countries with no data. Population statistics they used for the year 2005 is based on the United Nations Department of Economic and Social Affairs (UNDES 2011) database, while gridded global

population numbers are taken from the Columbia University Centre for International Earth Science Information Network (CIESIN) database. They used the population cohort of 30 years and older in the health impact function.

Applying the health impact function to the considered population, the DREAM model output suggests a significant contribution of desert dust to premature human mortality. For the global background of dust concentration of 7.5 µg/m3 i.e., threshold below which no premature mortality occurs, the estimated premature mortality (per grid cell) by cardiopulmonary disease (CPD) and lung cancer (LC) is illustrated in Figure 18. In total, around 570,000 premature deaths in the model domain are predicted to occur during a one-year period, as a negative consequence of dust. According to our results, top five countries with the highest induced CPD-mortality in the year 2005 are: Egypt with 74,000; Iraq with 67,000; Iran with 50,000; Nigeria with 46,000; Sudan with 45,000. On the other hand, top five countries with the highest induced LC-mortality in the same year are: Iraq with 1,200; Iran with 900; Sudan with 800; Egypt with 800; Uzbekistan and Turkey with 500 premature deaths each.

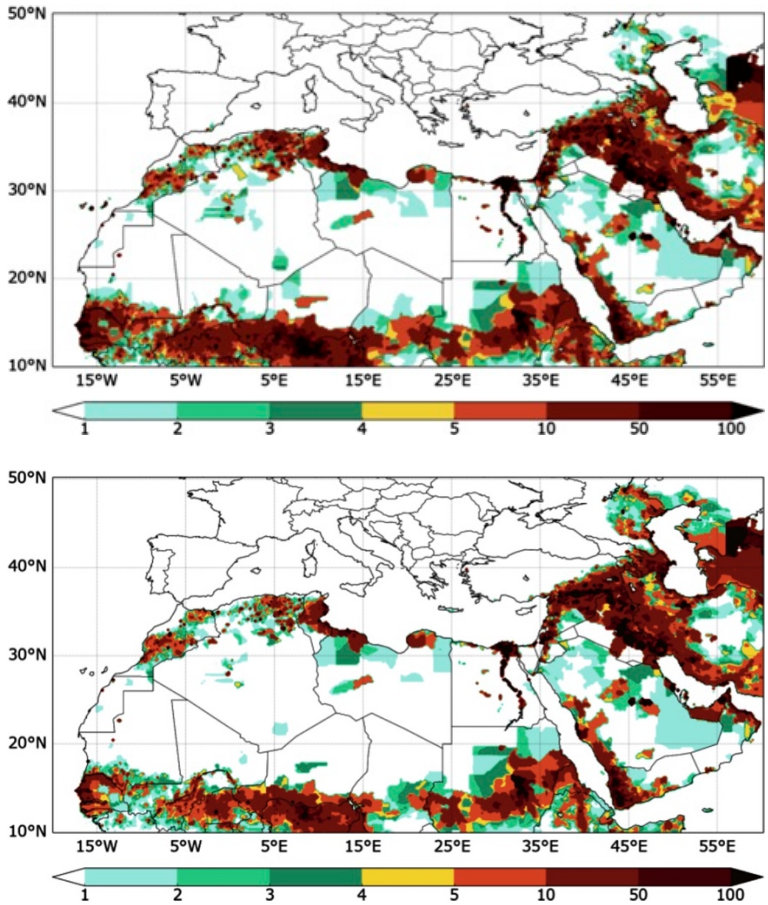


Figure 18 - Estimated global premature mortality per grid cell by cardiopulmonary disease (top) and lung cancer (down) due to the long-term exposure to desert dust with an aerodynamic diameter smaller than 2.5 µg/m³, calculated by the VI-SEEM DREAMCLIMATE service.

Researchers also investigated the sensitivity of their results on the value of the threshold concentrations, which is above assumed to be 7.5 μg/m3. This analysis is presented to showcase capabilities of the model and the developed DREAMCLIMATE service, and can be efficiently used to study desired regions and time periods if the required input data are published. Future work in this area will produce transatlantic results i.e. simulation will cover wider domain of 5,000 x 10,000 km with horizontal resolution of 15 km.

7.1.1.8 DACEM1

Dust Aerosols Climatic Effects over the Mediterranean - Phase 1
Climate research – Area A

The implementation of DACEM1 resulted to the production of a high-resolution atmospheric gridded dataset for the region of Europe and North Africa (2N – 60N, 20W – 62E). More specifically, the data were produced by the conduction of numerical simulations with the WRF-Chem model at a horizontal grid spacing of 0.11o x 0.11o (~12 x 12 km2). The data span the 30-year period from 1987 through 2016, representing the most recent climatology of the study area. Overall, the output of DACEM1 constitutes a prototype atmospheric dataset that can be exploited for supporting several climate-related applications and services. For instance, the dataset has been employed for providing a prototype climatology and trend analysis of the thermal bio-climate of the Euro-Mediterranean region. This climatology is one of the first of its kind, providing useful insights on the impacts of climate change on human thermal comfort and perception.

Scientists continuously update the original atmospheric gridded dataset, including most recent years and they work towards the development of climate-based applications focusing on tourism and agrometeorology.

7.1.1.9 AUTH_WRF371M_EUR0.44

AUTH contribution to EURO-CORDEX 0.44 regional climate projections
Climate research – Area A

High-resolution climatic information produced by regional climate models is essential for the investigation of the impacts of a changing climate occurring on regional and national scales. The impact of this application is to i) better understand regional climate phenomena, their variability and changes within the century, improving thus the knowledge on climate science and to ii) provide the scientific community with open-access climatic data set, which will be available as a EURO-CORDEX ensemble member.

This project is based on numerical computations of the WRF regional climate model. The simulations cover a time slice of twenty years in a grid of 106x103 grid cells and stored in 3 hourly values. The need for computational and storage resources are essential for this project.

In future work, performance of higher spatial resolution climate projections will cover the whole European domain until the end of the century.

The mean seasonal temperature of the WRF371 simulation for the period 1990-2008 can be viewed in Figure 19.

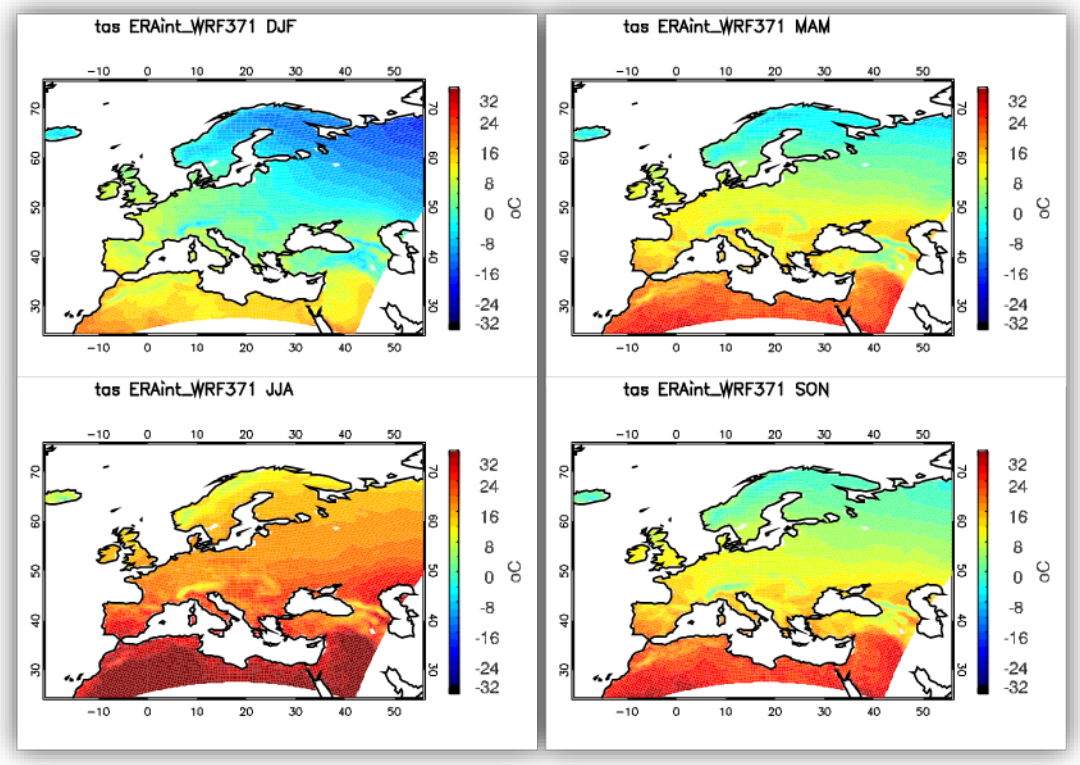


Figure 19 - Mean seasonal temperature of the WRF371 simulation for the period 1990-2008.

7.1.1.10 WRF-ARW

WRF-ARW: Weather forecasting with chemistry over Cyprus
Climate research – Area C

Researchers employed the WRF-Chem model to study summertime air pollution, the intense photochemical activity and their impact on air quality over the eastern Mediterranean. They utilized three nested domains with horizontal resolutions of 80, 16 and 4km, with the finest grid focusing on the island of Cyprus, where the CYPHEX campaign took place in July 2014.

Anthropogenic emissions are based on the EDGAR HTAP global emission inventory, while dust and biogenic emissions are calculated online. Three simulations utilizing the CBMZ-MOSAIC, MOZART-MOSAIC, and RADM2-MADE/SORGAM gas-phase and aerosol mechanisms are performed. The results are compared with measurements from a dense observational network of 14 ground stations in Cyprus. The model simulates T2 m, Psurf, and WD10 m accurately, with minor differences in WS10 m between model and observations at coastal and mountainous stations attributed to limitations in the representation of the complex topography in the model. It is shown that the south-eastern part of Cyprus is mostly affected by emissions from within the island, under the dominant (60%) westerly flow during summertime. Clean maritime air from the Mediterranean can reduce concentrations of local air pollutants over the region during westerlies. Ozone concentrations are overestimated by all three mechanisms (9% ≤ NMB ≤ 23%) with the smaller mean bias (4.25ppbV) obtained by the RADM2-MADE/SORGAM mechanism. Differences in ozone concentrations can be attributed to the VOC treatment by the three mechanisms. The diurnal variability of pollution and ozone precursors is not captured (hourly correlation coefficients for O3 ≤ 0.29). This might be attributed to the underestimation of NOx concentrations by local emissions by up to 50%. For the fine particulate matter (PM2.5), the lowest mean bias (9μgm−3) is obtained with the RADM2-MADE/SORGAM mechanism, with overestimates in sulphate and ammonium aerosols. Overestimation of sulphate aerosols by this mechanism may be linked to the SO2 oxidation in clouds. The MOSAIC aerosol mechanism overestimates PM2.5 concentrations by up to 22μgm−3 due to a more pronounced dust component compared to the other two mechanisms, mostly influenced by the dust inflow from the global model. Researchers conclude that all three mechanisms are very sensitive to boundary conditions from the global model for both gas-phase and aerosol pollutants, in particular dust and ozone. In future work, zero-out emissions test for the island of Cyprus and pollution transport in the East Mediterranean region will be included. An example of WRF-ARW model simulation over the Mediterranean can be found in Figure 20.

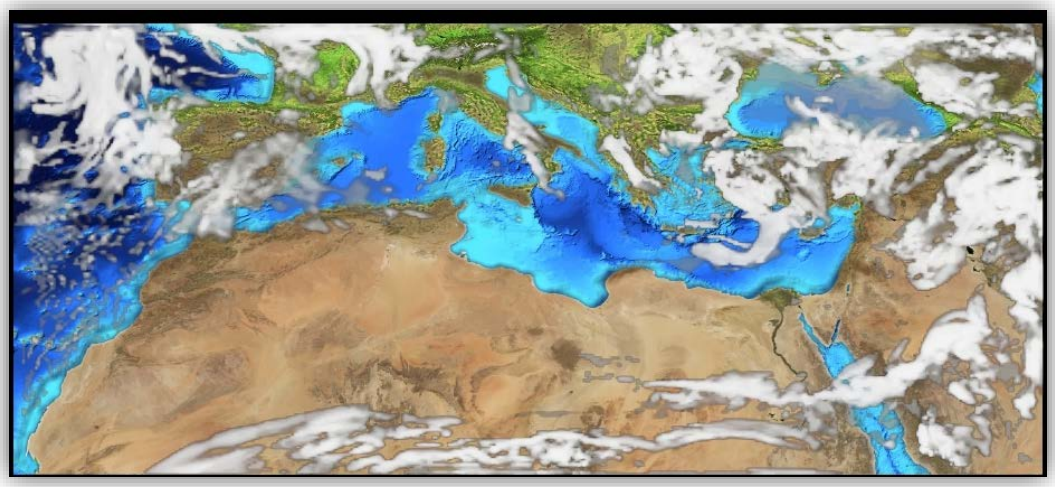


Figure 20 – WRF-ARW model simulation over the Mediterranean.

7.1.1.11 DACEM2

Dust Aerosols Climatic Effects over the Mediterranean - Phase 2
Climate research – Area A

Following the successful implementation of DACEM1, DACEM2 exploited the knowledge and experience gained in order to focus on specific aspects of the climate-aerosol system in the Mediterranean region. In particular, an international collaboration was established for studying dust aerosol effects on Mediterranean tropical-like cyclones (Medicanes). The collaboration was established in the frame of the HyMeX initiative and currently involves researchers from Greece, Italy, Cyprus and France. Using the HPC allocated time, high-resolution numerical simulations were conducted, focusing on a particular Medicane event that took place in December 2005. The goal is to examine in detail in the influence of dust aerosols on the structure, physics and dynamics of the Medicane. Further, numerical simulations were also carried out to study the direct impacts of dust aerosols on the radiation budget and, consequently, the meteorology of the south-east Mediterranean region.

Continue the international collaboration on dust aerosol impacts on Medicanes and finalize numerical experiments and data analysis. Conduct further experiments for investigating dust aerosols direct and indirect effects.

7.1.2 Applications in Life Sciences

7.1.2.1 PSOMI

Protein-Small Organic Molecules Interaction
Life Sciences - Area A

The researchers examined molecular dynamic simulations between series of new molecules with lysosome. The initial simulation results showed that there are no interactions between the synthesized molecules (enol-carbonates) and the lysosome.

By using computer simulations, which are characterized by high accuracy and precision, would provide valuable information during the research work in the chemical laboratory, in order to design new drugs.

On the one hand, it was possible to investigate the interactions of such molecules that have not been made in the laboratory, and in terms of their structural or physiological properties, they are similar to those existing in nature (or synthesized artificially). On the other hand, the application of simulations contributes to a significant reduction in research time, and orientation to those groups of molecules that are highly likely to exhibit a physiological effect after interaction with the given receptor. In this way, the number of molecules that are planned to be tested is significantly reduced.

Finally, by exchanging some functional groups with some other, researchers would get useful information about the possible interactions (and activities) of such molecules, which would significantly contribute to the synthetic application of the tested reaction or methodology.

In future work, there will be examination de novo interaction of an organic molecule with protein. Subsequently, it would be preferable to vary or branched proteins in place of lysosomes or other classes of synthetic organic molecules instead enol-carbonates. Pictures of investigated molecules are presented in Figures 21 - 24.

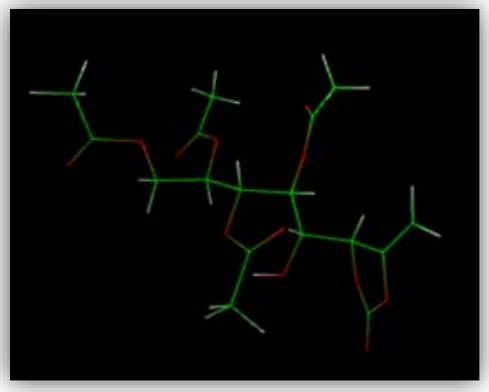


Figure 21 - (2R,3R,4R,5S)-5-hydroxy-5-((S)-5- methylene -2-oxo-1,3- dioxolan -4-yl)pentane-1,2,3,4-tetraacetate in program VegaZZ (PDB format)

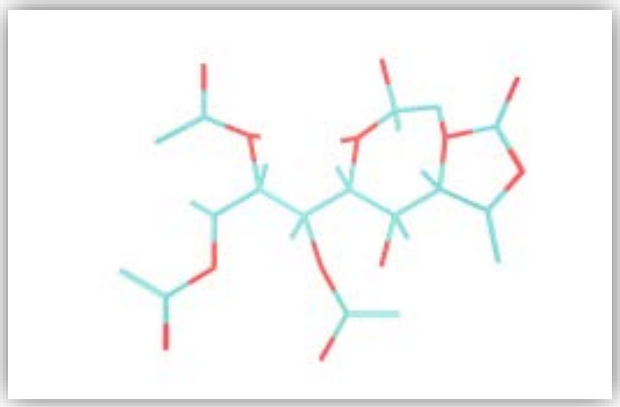


Figure 22 - (R)-4-((R)-hydroxyphenyl-methyl -5-methylene-1.3-dioxolan-2-one

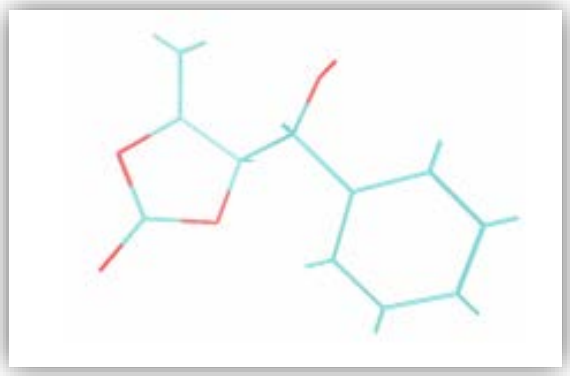


Figure 23 – (R)-4-((R)-(4-chlorophenyl)hydroxyl methyl-5-methylene-1.3-dioxolan-2-one

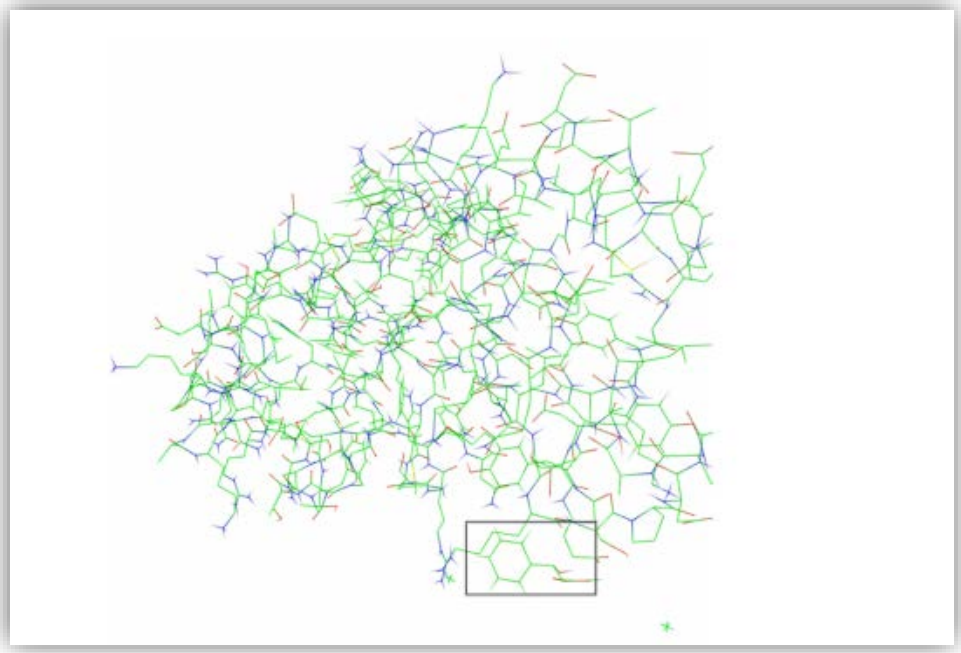


Figure 24 – The most stable possible conformation of protein and ligand (VegaZZ)

7.1.2.2 MULTIDRUG

Multiscale computations to reveal the potential of molecular drug carriers
Life Sciences – Area A

Based on an extensive survey of the scientific literature, the folate receptor-α (FR-α) was chosen as the target cell surface protein to be recognized by a steering vector, part of a drug delivery system (DDS). Folic acid (FA; the natural ligand of FR-α) and some of its analogues (MTX, RTX, PTX, MTHF, PON; termed antifolates) were chosen to play the role of targeting molecules. The first step from the investigation of such complex DDS is the characterization of the structure of the ligand-vectors. Atomistic classical molecular dynamics simulations were performed within the application MULTIDRUG to gain knowledge of their behaviour in conditions that mimic the biological environment - aqueous solution, ionic strength, body temperature, and ambient pressure. Four of the ligands (FA, MTX, MTHF, PON) were also modelled in trans and cis configuration along an amide bond and the other two (RTX and PTX) – only in trans conformation. This resulted in 10 independent trajectories. Using standard MD protocol, 100 ns of each trajectory were generated with Gromacs 5.1.2 on Avitohol. The last 50 ns were subjected to statistical analysis.

The structural dynamics of the trans and cis forms was characterized. This included estimation and analysis of root-mean-square deviation (RMSD) profiles, distribution of the head-to-tail distance, cluster analysis, hydrogen bonding, various quantum chemical calculations. It was found that the trans forms are more mobile and the cis ones – more rigid. The cis forms could be populated spontaneously for FA and PON. Overall, the molecules of the six ligands were classified as structurally mobile. The structural flexibility of trans MTHF was the most pronounced. Different preferred shapes were determined for the various ligands, too. This implies dissimilar behaviour of the six ligands and of the two configurational isomers with respect to receptor binding.

A second part of the simulations within MULTIDRUG were those of a mixed lipid bilayer, which mimicked part of a cancer cell membrane, with one molecule of the folate receptor-α embedded into it. MD simulations of this system in aqueous solution and in the presence of NaCl, as in physiological conditions, were performed. The total size of the model was ca. 185000 atoms. The effect of the pressure scaling on the behaviour of the model membrane and of the receptor was checked by applying two different algorithms - an isotropic and a semi-isotropic one. Hence, two independent MD trajectories were generated, each of length 150 ns. The last 50 ns were subjected to statistical analysis.

The properties of the model systems were assessed by generating mass density profiles, order parameters of the lipid tails, areas per lipid molecule. The secondary structure of the protein was also analyzed. It was found that the pressure scaling scheme had an effect on the structure of the lipid bilayer - the isotropic scaling maintained liquid phase state, while the semi-isotropic one induced more order of the lipid molecules, which is closer to the experimental findings. The spatial structure of the protein molecule was only slightly affected. The other main focus of the work within the application MULTIDRUG was to facilitate the passage of drug substances through cell membrane models. For this purpose, a DDS component (DOX-CPP) was composed of a peptide passing through the lipid bilayer (CPP) covalently bound to a molecule of the chemotherapeutic doxorubicin (DOX). Two types of membranes, each of 162 lipids, one containing only DPPC molecules and the other constructed

of six lipid types mixed in a proportion similar to that in the erythrocyte cell membrane, were studied. The bilayers were hydrated on both sides and simulated both in the presence and in the absence of the complex DOX-CPP. The size of the models amounted to ca. 45000 atoms. Atomistic classical MD simulations in physiological conditions were performed with two pressure scaling methods (isotropic and semi-isotropic) and two types of periodic boundary conditions (two-dimensional and three-dimensional). The productive part of each simulation was 1 μs. 10 independent MD trajectories were generated. The statistical analysis of the MD trajectory showed that the pressure scaling method had a significant effect on the order parameter of lipid tails, lipid surface area, and lateral diffusion coefficients. The type of boundary conditions influenced significantly the lateral diffusion coefficient and the order parameter. These descriptors are an indication of the phase state of the membranes and determine the entry of DDS into the lipids. Better DOX-CPP penetration was achieved with the isotropic pressure scaling with three-dimensional periodic boundary conditions and this method was more appropriate for the purposes of the study. With these models, three umbrella sampling simulations were performed. They consisted of a total of 120 MD trajectories, each of 10 ns length. The free energy of transfer of DOX-CPP through the membranes was calculated. Based on these results, a longer 3 μs simulation was performed with another position of the complex to the membrane. The peptide penetrates the membrane halfway and provides information about a possible passage mechanism.

The outcome of the simulations improved the molecular-level understanding of how drug-peptide complexes penetrate through cell membranes. This should facilitate the more efficient design of cell-penetrating entities for better drug transfer into cells. Some of the calculations provided the necessary reference properties for further studies of the mechanism of directed drug transport. This is the major direction of the future computations - understanding how the vector ligands are attracted by the surface receptor embedded into the model membrane and whether they can carry drug-peptide complexes as a bioactive cargo.

All the calculations described above would have been impossible within this time period without the provided computational resources. The allocated core hours on the HPC facility enabled efficient outcome of results on biochemically reasonable models.

The MD simulations will be extended to models where the interactions between a vector-cargo conjugate and a membrane-embedded folate receptor-α are described. Comparison between the behaviour of the ligands with and without attached drug payload will be made. Some quantum chemical calculations on partial models of the protein active centre will also be needed to quantify the drug-receptor binding.

Figure 25 shows the initial structure for the MD simulations of the receptor-embedded lipid bilayer. Figure 26 is a 3D representation of the secondary structure of the folate receptor-α after 150 ns of MD simulation with semi-isotropic pressure scaling. Figure 27 is a representation of the most preferred shapes of the vector ligands after 100 ns of MD simulation. Figure 28 provides a snapshot where the complex DOX-CPP has penetrated the lipid bilayer to the largest extent.

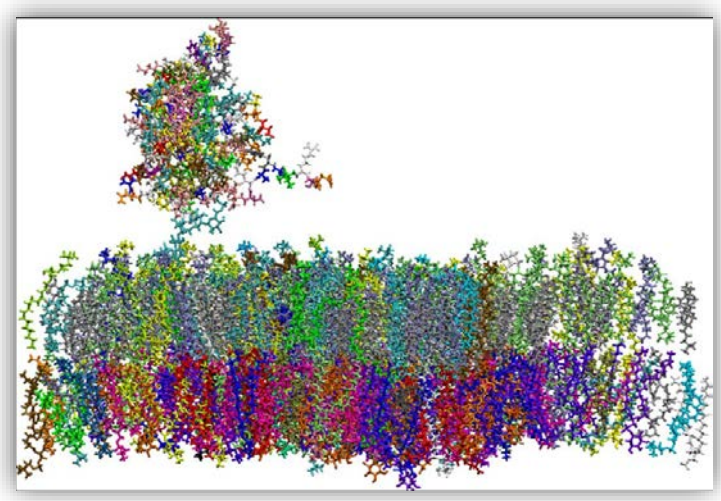


Figure 25 - The initial structure for the MD simulations of the receptor-embedded lipid bilayer

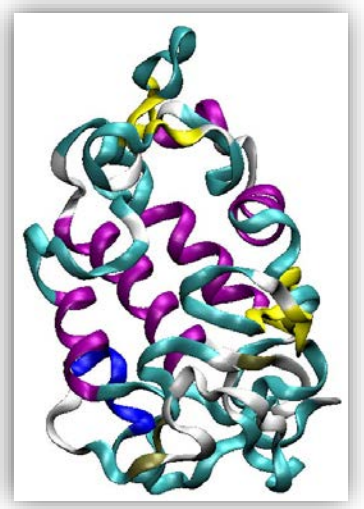


Figure 26 - 3D representation of the secondary structure of the folate receptor-α after 150 ns of MD simulation with semi-isotropic pressure scaling.

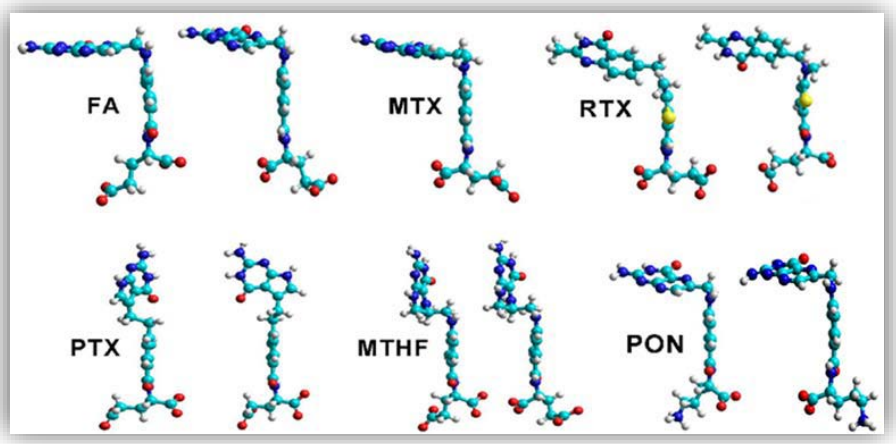


Figure 27 - A representation of the most preferred shapes of the vector ligands after 100 ns of MD simulation

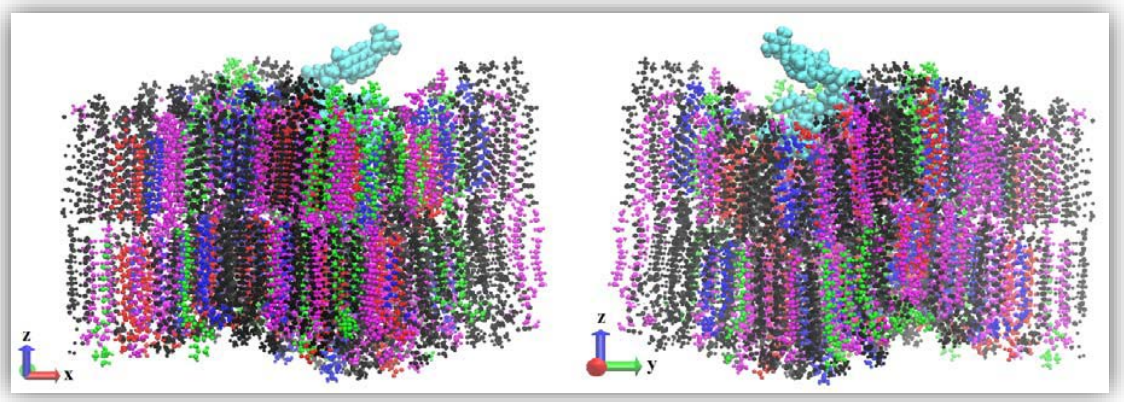


Figure 28 - A snapshot where the complex DOX-CPP has penetrated the lipid bilayer to the largest extent

7.1.2.3 D3R

Predicting Farnesoid X receptor (FXR) - inhibitor structures and affinities for computer-aided drug design within the D3R challenge
Life Sciences – Area B

At the beginning of the challenge researchers were provided with the structures of 36 ligands which are known to bind in FXR, so as to predict their binding modes. Two subsets of 15 and 18 compounds which belong in the same congeneric series, and therefore are rendered suitable for relative binding free energy exercises, were also given to us for the free energy

predictions part. Researchers chose to follow two different methodologies for the compounds of the pose predictions according to whether there were already solved crystal structures in the Protein Data Bank with ligands of same chemotypes, or whether the ligands of the dataset were entirely new. In the first case the main methodology consisted of (a) docking in the structure with the chemically similar ligand, (b) alignment to it and (c) minimization of the complex. Compound FXR_9, which is a case where it was used this methodology, was initially docked into a structure with a chemically similar native ligand, then it was aligned to the native ligand and finally the whole system underwent energy minimization (Figure 29).

In the second case, where the chemotypes of the ligands were unknown, researchers chose which crystal structure to dock the ligands into, based on shape similarity, cross docking and interaction similarity calculations. These two methodologies were enhanced, where needed, by examination of water thermodynamics, binding pose metadynamics and free energy calculations. In the case of FXR_6 compound (Figure 2), the ortho methoxy substituent was not unambiguous which of the two equivalent ortho positions should occupy. Therefore, the researchers perform water thermodynamics calculations with the tool WaterMap of the Schrödinger Suite, which predicts regions in the binding pocket that are probable to be occupied with thermodynamically stable of unstable water molecules. According to this analysis, in Figure 30 is shown that if the ortho methoxy is on the left of the ring then it may replace an unstable water molecule. Thus, this pose was submitted as the most probable one. Another case where the place of an ortho substituent was not clear, is that of FXR_27, which has an ortho chloro substituent. For this compound they performed Free Energy Perturbation (FEP) calculations, which indicated the upward placement of the substituent as the most stable (Figure 31).

For the ligands with the known chemotypes, they predicted quite accurately their true binding modes, while for those with the unknown chemotypes, our protocol performed well in predicting a specific group’s new binding mode, but there is still room for improvement considering its whole performance. In total, this research team ranked 1st out of 49 completed submissions, with a median RMSD value of 0.99 Å. These results validate the adopted methodology, especially for compounds with known chemotype.

As for the relative binding free energies part, the team performed Free Energy Perturbation calculations coupled with Molecular Dynamics simulations for both subsets. At the end of the first stage of the challenge, the true binding modes of the two sets were released and so they repeated the calculations considering the compounds’ crystal poses this time. The results for both subsets are presented in Figure 32. For the spiros subset, a significant improvement of the free energy predictions was observed after the release of the crystal structure, whereas for the sulfonamides set, the range of the experimental binding affinities was within the error of the method, thus not allowing for accurate predictions. It is encouraging, though, that in the 2nd stage of the challenge all compounds were predicted as less potent than the reference molecule (green square box inside a circle), and therefore they would not have been proposed for organic synthesis, thus saving valuable resources in a lead optimization project.

The specific project was completed with the end of D3R Grand Challenge 2. However, Grand Challenge 2 was part of a series of blinded challenges, together with SAMPL and CELPP, also organized by D3R. These challenges motivate the ongoing effort for the generation of new and efficient methodologies, as well as algorithms, related to the accurate prediction of binding modes and affinities between ligands and protein-targets of pharmaceutical interest. Concluding, computational methods in drug design have shown significant advances over the years, and future developments are expected to lead to continued improvement in performance in blinded prediction challenges.

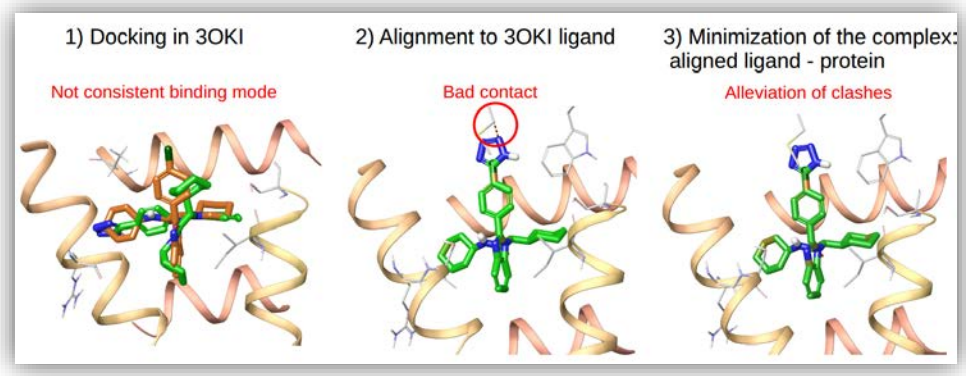


Figure 29 – Consecutive steps for the prediction of FXR_9 (green) binding mode. 1) Docking in 3OKI crystal structure which has a native ligand of the same chemotype (orange). 2) Alignment of FXR_9 in the native ligand. 3) Energy minimization of the FXR_9 – protein complex.

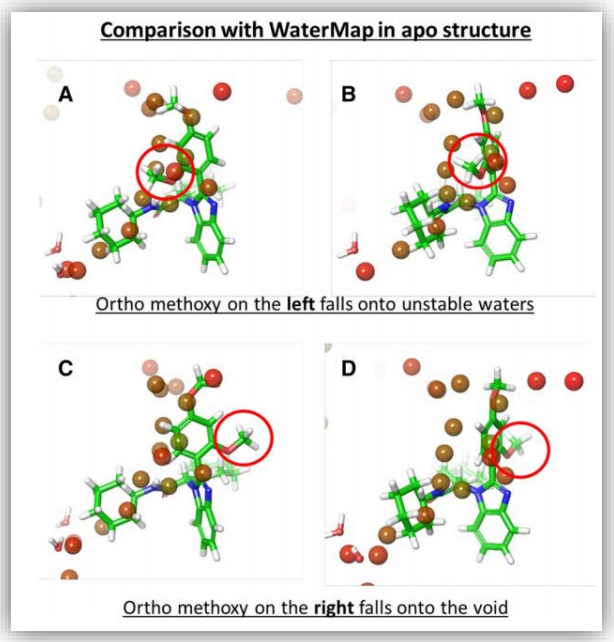


Figure 30 – Superposition of FXR_6 possible methyl orientations with WaterMap calculation in the apo structure. a, b Show the conformation of FXR_6 when the methyl substituent is on the left of the ring and c, d when it is on the right.

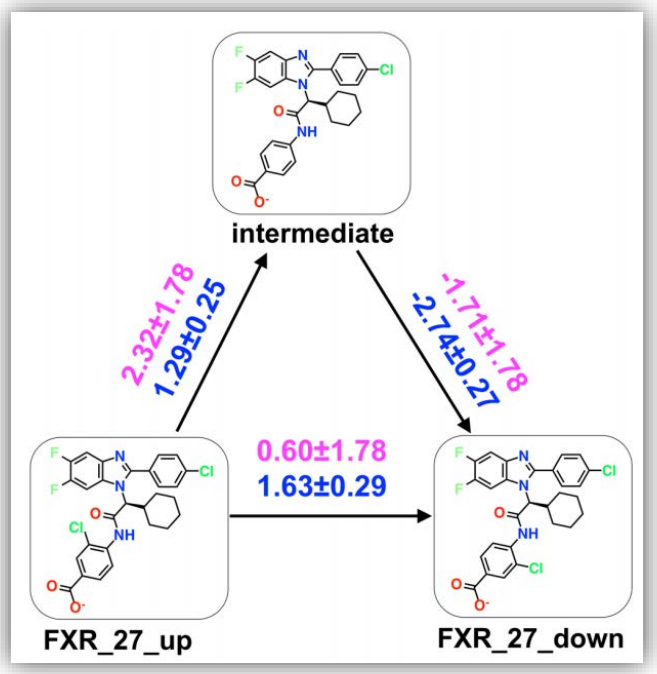


Figure 31 – FEP results for the identification of the most stable pose of the ortho substituent for FXR_27. The upward placement of the chloro substituent is more stable by 0.60 ± 1.78 kcal/mol.

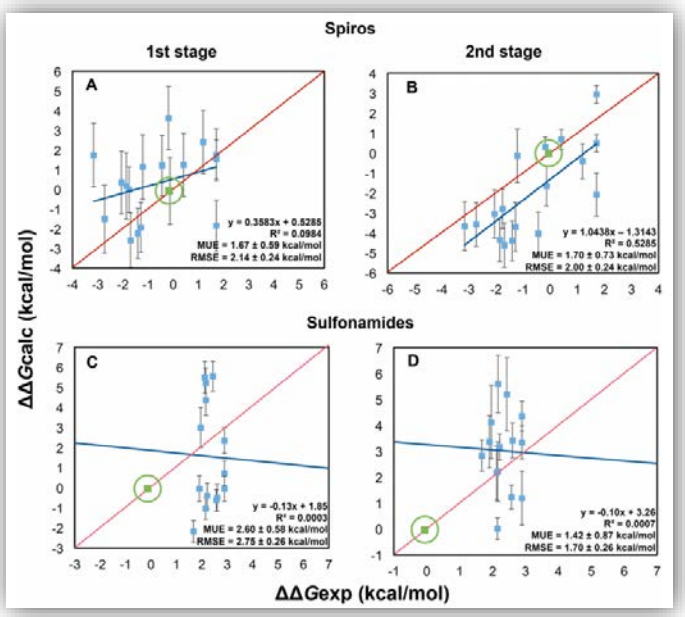


Figure 32 – Correlation diagrams of the predicted and the experimental free energies for spiros (a, b) and sulfonamides (c, d) compounds in both stages of the challenge. The y=x line is also shown in red. The point (0, 0), which corresponds to the reference compound, is shown in green square box inside a circle.

7.1.2.4 MolSurf

Molecular description of the rheological properties of natural and synthetic surfactants at the air-water interface.
Life Sciences – Area A

In order to study in more detail the alignment and the interactions between the molecules of the saponin surfactant escin in dense adsorption layers at the gas/water interface, long atomistic molecular dynamics (MD) simulations of a dense layer of neutral escin adsorbed on the surface of a water slab were carried out. The accomplished length of the MD trajectory was 1 μs (generated as 20 trajectory parts, 50 ns each) for a model system with 441 molecules Escin in water composed of more than 43000 molecules. The total size of the model was 203139 atoms.

The analyses of the number of hydrogen bonds within and between escin molecules over time showed that there was a trend for decrease of the number of H-bonds when the simulation progressed. This indicated ongoing change of the intermolecular orientation and required the above-mentioned long simulation time. After ca. 350 ns the number of H-bonds reached a constant average value. This is then the characteristic relaxation time for a model of this size (441 molecules). This time is longer than that for the small models obtained in our previous study. The rest of the simulation up to 1 μs was performed to collect data for reliable statistical analysis, which is now ongoing.

Extensive data processing is currently done to extract results about the surfactants orientation at the interface and their characteristic sizes and degree of immersion of escin in the water. Adsorption layer thickness and surface tension is also being estimated and will be compared to experimentally obtained values to assess the model. To check the hypothesis that the substantial elasticity of neutral escin layers is explained with the presence of a specific interaction, which is intermediate between hydrogen bonding and dipole-dipole attraction, some specific analyses will be made, too.

The obtained results will provide more insight into the behaviour of escin adsorption layers on the water surface, which will enable their more efficient application in different washing formulations. The use of escin is of interest to the industry because it is a more eco-friendly alternative to the conventional synthetic surfactants because of its natural origin and the possibility for biodegradation.

The availability of the allocated computational resources within the application MolSurf was key to performing the research described above. The chance to construct and simulate this much larger and much more realistic model of an escin layer enabled the verification of the conclusions made at the smaller scale and the observation of collective behaviour of the surfactant molecules, e.g. well expressed curvature of the surface in both lateral dimensions

that would have been inaccessible with smaller models. The achieved microsecond time scale also allowed the characterization of the intermolecular reorientation events taking place on a longer time scale.

The plans for future work are to investigate the surface elastic modulus for Escin adsorption layers with non-equilibrium molecular dynamic simulations.

7.1.2.5 CNCADD

Conventional vs Novel Computer Assisted Drug Delivery
Life Sciences – Area A

Hydrophilic drug irinotecan and morphine (in both neutral and protonated cationic form for the later system) studies have been thoroughly performed, from various aspects, with main general aim, the development of a hybrid statistical physics – quantum mechanical approach which would enable the understanding of interactions between drug molecules and polymeric nanocarriers on a fundamental (molecular) level, i.e. to go far beyond the phenomenological descriptions in this area. Such understanding is essential in the course of development of new drug nanocarriers as well as for in-depth insights into the molecular-level mechanisms that govern the drug inclusion processes. In the first step, the initial computational studies of irinotecan involved “static” exploration of the potential energy surface of this molecule at several semiempirical and density functional (DFT) levels of theory, followed by computation of its vibrational spectra. These computational studies have been done with main aim to understand how the interaction with certain drug nanocarriers modifies the intramolecular vibrational force field, and how can one judge on the intermolecular interactions of this type by analyzing the vibrational force field. However, essentially all of the relevant processes in living organisms take place at finite temperatures, quite above absolute zero, and the static computations actually refer to free molecular species at absolute zero temperature. To account for the system’s behaviour under realistic conditions (e.g. to simulate the vibrational spectra of this molecule at various finite temperatures), the researchers have also undertaken molecular dynamics simulations thereof. These simulations have been carried out employing the atom-centred density matrix propagation (ADMP) scheme. They have further on compared the dynamical results with the ones from the “static” studies. Studies of the temperature-dependent evolution of the spectra have provided rather useful structural insights for this system.

A robust computational approach has been also implemented towards a better understanding of solid-state infrared spectra of morphine-based formulations. The approach seems to contribute substantially to a clarification of the issue related to the forms in which morphine molecular system packs within solid state formulations thereof. In this context, solid state vibrational spectroscopic properties of morphine (in the form of morphine sulphate pentahydrate) were studied combining experimental and theoretical approaches. Experimental studies involved a Fourier transform infrared spectroscopic (FT IR) study of the

title compound with the attenuated total reflection (ATR) technique. From theoretical side, a detailed study of the main possible constituents of the investigated molecular crystals: neutral morphine molecule and N-protonated morphine-H+ cation was carried out, employing meta hybrid density functional theory (DFT) approach with the M06-2X exchange-correlation functional, using 6-311G(d,p) basis set. Potential energy surfaces of the studied systems were explored in details, paying particular attention to the intramolecular torsional flexibility. In both cases, four stable minima on the explored PES were located. Subsequently, harmonic vibrational analyses were carried out for each minimum, with the method of diagonalization of the mass-weighted Hessian matrices. Experimental data strongly suggest the presence of hydrogen sulphate anionic species in the investigated solid phase. It is suggested that both neutral (unprotonated) morphine along with the N-protonated morphine-H+ cation could be present in the solid state as well. The agreement between experimental and theoretical spectra in the regions of appearance of bands due to morphine (or N-protonated morphine-H+ cation) intramolecular vibrational modes is remarkable. Thus, the present theoretical calculations enable solid theoretical support to the empirical assignments of the spectral bands and it could serve as a starting point for further studies of spectroscopic manifestations of morphine incorporation into nanocarrier drug delivery systems.
Further on, the research team have also studied the kinetics of morphine release from various solid-state formulations (produced by both optimized technological processes as well as from the screening phase of formulation development). This was done by various macroscopic kinetic models, which account for the complexity and the multistep nature of drug release from the particular formulations. The implemented macroscopic models have, up to this point, allowed to get substantially relevant insights into the mechanisms governing drug release from various formulations, as well as a molecular-level understanding of the processes taking place in the course of these phenomena.

In the course of the last activity, CNCADD is currently aiming at the development of a generalized model that would relate the molecular-level processes with the macroscopically observed kinetics of drug release under in vitro conditions.

There have been also performed more detailed and more in-depth studies of free irinotecan molecule by more systematic dynamical approach. These studies involved a series of molecular dynamics simulations of this molecule with the atom-centered density matrix propagation scheme at AM1 semiempirical level of theory, at series of different temperatures, ranging from 5 K to 300 K. Molecular dynamics simulations were performed within the NVE ensemble, initially injecting (and redistributing among the nuclei) various amounts of nuclear kinetic energies to achieve the desired target temperatures. Subsequently to initial equilibration phase of 2 ps, productive simulations were carried out for 8 ps. The accuracy of simulations and the closeness of the generated trajectory to those at the Born-Oppenheimer surface were carefully followed and analyzed. To compute the temperature-dependent rovibrational density of states spectra, the velocity-velocity autocorrelation functions were computed and Fourier-transformed. Fourier-transformed dipole moment autocorrelation functions were, on the other hand, used to calculate the temperature-dependent infrared absorption cross section spectra. The finite-temperature spectra were compared to those

computed by a static approach, i.e. by diagonalization of mass-weighted Hessian matrices at the minima located on the potential energy surfaces. Thermally-induced spectral changes were analyzed and discussed. The advantages of finite-temperature statistical physics simulations based on semiempirical Hamiltonian over the static semiempirical ones in the case of complex, physiologically active molecular systems relevant to intermolecular interactions between drugs and drug carriers were pointed out and discussed.

The results obtained within the project, open certain new directions in the field of theoretical and computational studies of active drug components incorporated in various matrices, particularly nanocarriers. The research started in the present project is expected to be continued along the following directions:

- implementation and application of the methodologies developed and tested in the present project to a wider range of systems relevant to development of new drug delivery systems and to a more in-depth understanding of the drug – carrier interactions on a fundamental molecular level; in this context, special attention will be paid to studies of drug molecules incorporated in confined environments (nanopools within the drug nanocarrier systems);
- further development and adaptation of the numerical methods developed in the current project to understand the macroscopic kinetics of drug release from the carriers; in particular, the multiphase processes will be considered in more details in this context;
- extension of the current methodology for an in-depth understanding of the stability and reactivity of drug molecules incorporated in conventional and novel drug delivery systems;
- improvement of the existing and development or adaptation of new time-series analytic methods for more efficient and more detailed extraction of data from molecular dynamics trajectories;
- linking the molecular-level data to macroscopic models describing the kinetics of the drug delivery processes.

The following pictures demonstrating the following. Figure 33: The minimum located on the AM1 PES of irinotecan molecule. Figure 34: The kinetic energy spectra obtained by Fourier transformation of the velocity-velocity autocorrelation function for the series of simulations carried out for free irinotecan molecule at the four different temperatures in the frequency region from 500 to 4000 cm^{-1}. Figure 35: The dipole moment autocorrelation spectra obtained by Fourier transformation of the dipole moment autocorrelation function for the series of simulations carried out for free irinotecan molecule at the four different temperatures in the frequency region from 500 to 4000 cm^{-1}. Figure 36: The true minima located on the M06-2X/6-311G(d,p) PES of free morphine molecule. Figure 37 – The true minima located on the M06-2X/6-311G(d,p) PES of free N-protonated morphine-H+ molecule. Finally, Figure 38 demonstrates the theoretical infrared spectra of the four conformers of the free neutral morphine molecule calculated at the M06-2X/6-311G(d,p) level of theory.

Figure 33 – The minimum located on the AM1 PES of irinotecan molecule (the starting geometry for the ADMP simulations)

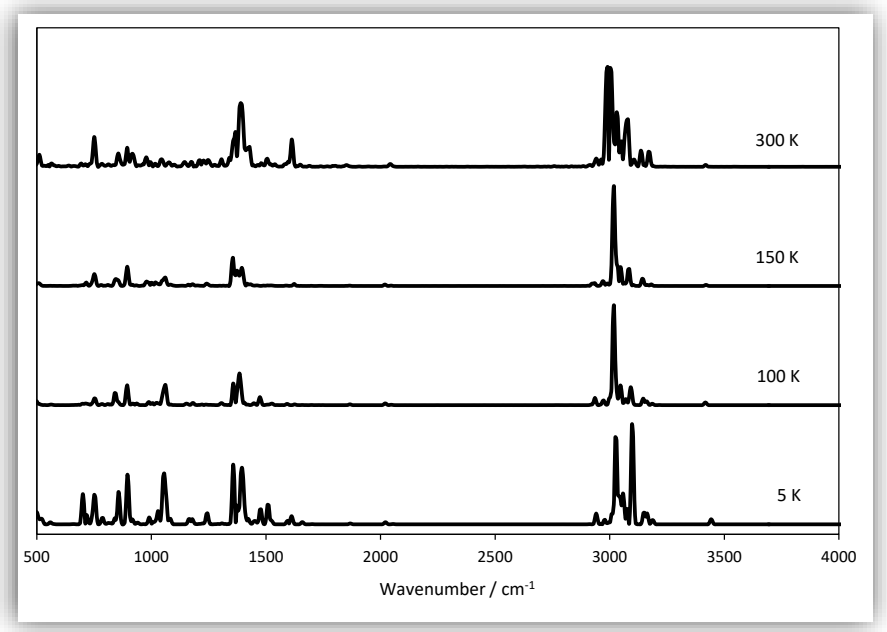


Figure 34 – The kinetic energy spectra (kinetic energy density of states spectra) obtained by Fourier transformation of the velocity-velocity autocorrelation function for the series of simulations carried out for free irinotecan molecule at the four different temperatures in the frequency region from 500 to 4000 cm^{-1}

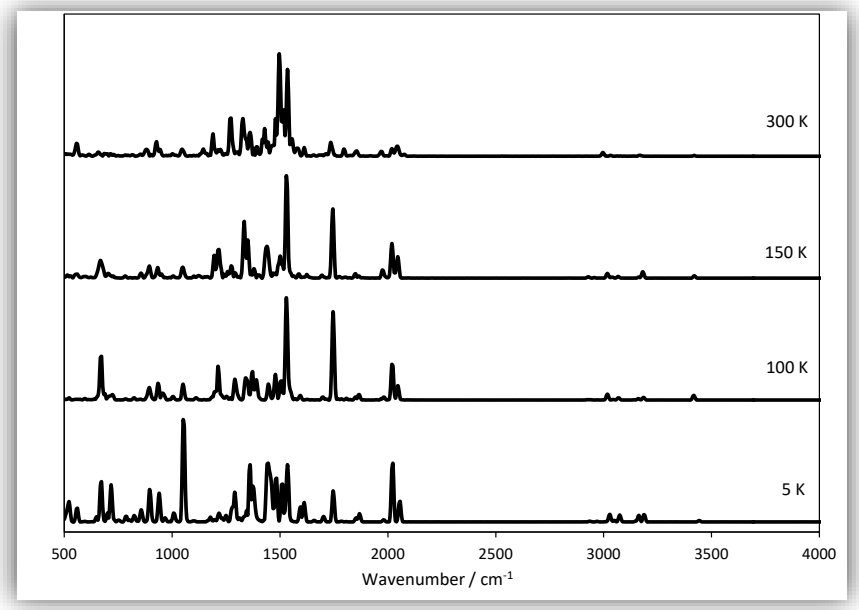


Figure 35 - The dipole moment autocorrelation spectra (~ infrared absorption spectra) obtained by Fourier transformation of the dipole moment autocorrelation function for the series of simulations carried out for free irinotecan molecule at the four different temperatures in the frequency region from 500 to 4000 cm^{-1}.

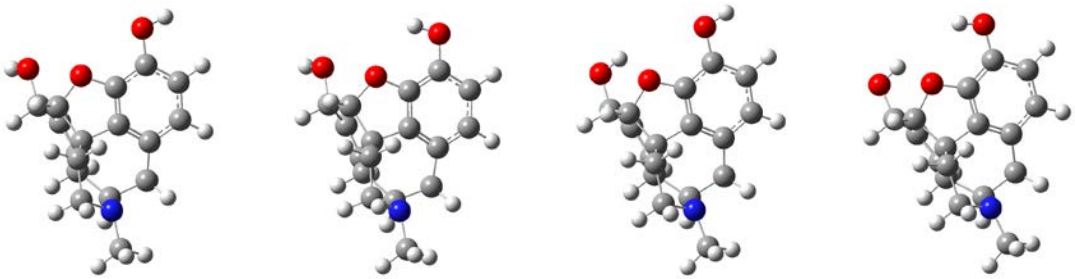


Figure 36 – The true (real) minima located on the M06-2X/6-311G(d,p) PES of free morphine molecule.

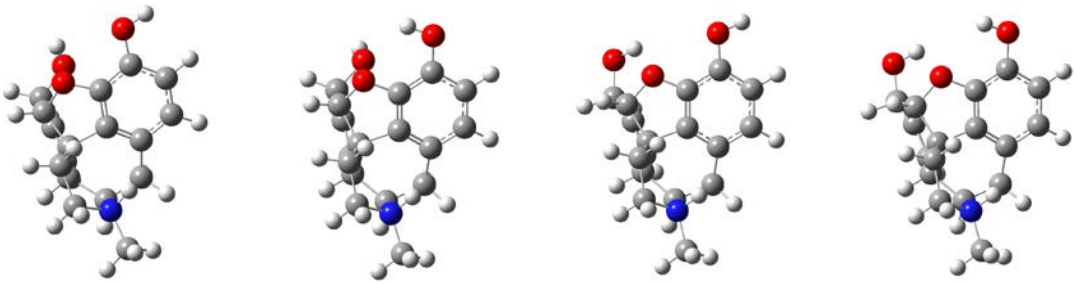


Figure 37 – The true (real) minima located on the M06-2X/6-311G(d,p) PES of free N-protonated morphine-H+ molecule.

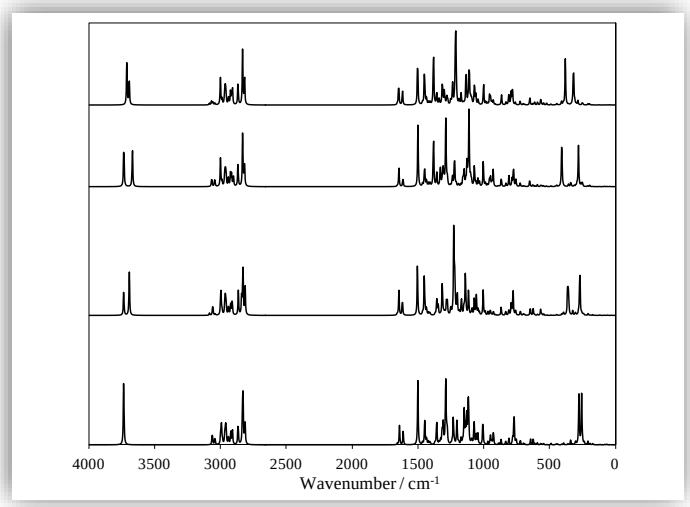


Figure 38 – Theoretical infrared spectra of the four conformers of the free neutral morphine molecule calculated at the M06-2X/6-311G(d,p) level of theory.

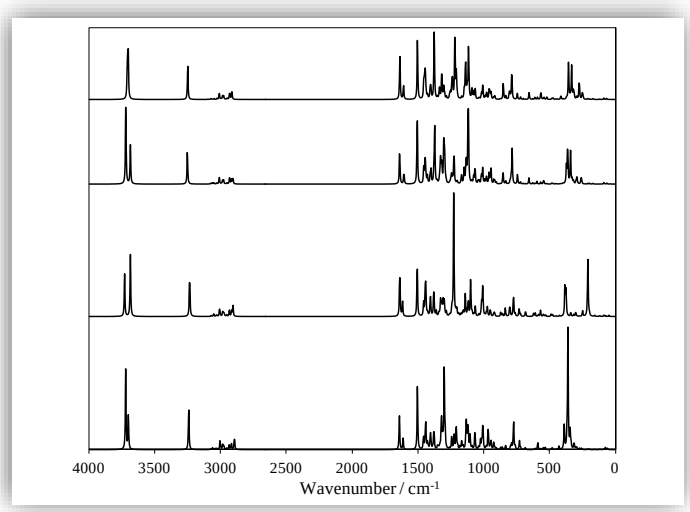


Figure 39 - Theoretical infrared spectra of the four conformers of the N-protonated morphine-H+ molecule (a, b, c and d in Fig. 3) calculated at the M06-2X/6-311G(d,p) level of theory.

7.1.3 Applications in Digital Cultural Heritage

7.1.3.1 CH-CBIR

Content-Based Image Retrieval and Classification in Cultural Heritage Applications Digital Cultural Heritage – Area C

In this project researchers showed that convnets show good results for remote sensing image classification and can be used as effective feature extractors even when trained on different datasets of remote sensing images. This suggests that the learned feature filters are universal to some extent. They also showed that convolutional feature extraction increases class separability compared to raw image pixels, and learning feature extraction filters further improve it.

Although there are no large labelled datasets of remote sensing images one can still apply convnets to high-resolution remote sensing (HRRS) image classification by means of transfer learning. Transfer learning is based on pre-training a convnet on a different dataset large enough for successful training and then fine-tuning it to the task manually. In order to assess the applicability of transfer learning to HRRS image classification, they used convnets pre-trained on ImageNet dataset for object recognition, and fine-tuned it on a dataset of HRRS images. They found out that both pre-trained and fine-tuned convnets are effective for feature extraction, as well as end-to-end HRRS image classification. They also analyzed the quality of the features extracted from various layers of the network from the standpoint of classification accuracy and showed that these features can be successfully used for image classification tasks. Moreover, convnets fine-tuned for HRRS image classification show good generalization abilities and can be used for classification of HRRS images obtained using different sensors, resolutions and imaging conditions without further fine-tuning. These results show that it is possible to train effective HRRS image classifiers using limited amount of training data, provided that there exists ample training images that share some common features with HRRS images. In this case the common feature is that both the ImageNet and HRRS images are obtained using visible spectrum.

The researchers developed an experimental framework for training and testing different convnet topologies for HRRS image classification. They also developed a web based interface for remote sensing image classification and made it available at http://mis.etfbl.net/cbir/. The classification results can be obtained as a web page, but also in JSON format, thus making the application ready for usage as a web service. The applications are developed using Python and libraries: Keras, Theano, Numpy, Scipy, Scikit-image, Scikit-learn, Matplotlib, and CherryPy. Source code is made available under MIT License in VI-SEEM code repository. Fine-tuned convnet models are also made available for other researchers and practitioners at Clowder.

Following the successful experiments with fine-tuning convolutional neural networks for classification of high-resolution remote sensing images, researchers turned their attention to

the problem of training convolutional neural networks using small number of manually labelled images. For example, in remote sensing there is a great body of multispectral, hyperspectral or radar imagery. Since there are no large labelled datasets of multispectral or hyperspectral remote sensing images, in order to use transfer learning they need large labelled datasets of such images in some other application domain. Unfortunately, large labelled datasets of e.g. multispectral or hyperspectral images, do not exist in any domain. Therefore they turned their attention to investigation of methods that would enable using of convnets for image classification in cases where scarce labelled data are available. One approach for solving this problem is to use self-supervised training. Self-supervised training is based on training a convnet on a surrogate task for which a large amount of training data can be cheaply obtained. The convnet trained on a surrogate task can be used for feature extraction or subsequently fine-tuned for the task manually with limited training data. In this project they have investigated image colorization and filling of missing spectral information as surrogate tasks. More specifically researchers have experimented with the approach known as split-brain auto-encoder and obtained promising results in HRRS image classification. Further investigation is needed to close the gap between the results obtained using self-supervised and traditional transfer learning, but the results obtained so far are an excellent starting point for further work.

Development of effective and efficient methods for HRRS image classification makes possible timely analysis of large body of remote sensing imagery that is being acquired on a daily basis. Availability of convnet models fine-tuned for HRRS image classification would make possible larger deployment in different application scenarios, where lack of available training data or computing resources for convnet training are obstacles for progress. Some of the application areas of HRRS image classification in cultural heritage that can benefit from this research are: detection of archaeological structures, monitoring of cultural heritage sites, detection of damage in non-accessible areas, assessment of environmental risks, etc.

Access to the computing infrastructure, especially GPUs is essential for research and application of deep learning. The most time and resource consuming part of this research is training of deep networks on large datasets of remote sensing images. In addition, training of deep networks requires hyperparameter tuning, meaning that it is necessary to repeat training multiple times with different hyperparameter settings. Using GPUs for training convnets enabled faster training and thus more exhaustive search for optimal hyperparameter values resulting in better classification accuracies and better applicability of the network for feature extraction.

The main body of work on HRRS image classification using convnets has been focused on images obtained using light in the visible spectrum. However, there are other imaging modalities of interest in remote sensing, such as multispectral and hyperspectral imaging, synthetic aperture radar (SAR) and Lidar, to name a few. In the future more work is needed in order to come up with effective and efficient classification approaches for images in other modalities.

One of the major obstacles for progress in remote sensing image classification is lack of large labelled datasets in modalities different than the visible spectrum. It is believed that self-supervised learning is an approach that holds promise to remedy this problem. However, it is necessary to investigate various surrogate tasks, convnet topologies as well as training procedures, and assess their ability for training convnets that would be useful for HRRS image classification.

Another important problem in application of convnets to remote sensing image classification is the size of images. State-of-the-art convnets in other application domains are typically trained on images whose dimensions are a few hundred pixels in both width and height. Remote sensing images can be a lot larger, a few thousand of pixels in width and height. Although conceptually convnets can be applied to large images, memory requirements for their processing prohibit direct application of existing convnet topologies and training procedures. It is necessary to find techniques for large-scale remote sensing image classification using convnets and our research interests are aligned with this goal.

Finally, with emergence of new sensing platforms, such as UAVs, the amount of imagery has started to grow at an unprecedented rate and the users of products based on remote sensing imagery come from very different areas. In some cases, it is necessary to analyze images on the sensing platform itself. However, the computing resources available on the sensing platforms are very limited in processing power, available memory and power consumption. Therefore, the researchers are interested in devising efficient convnet topologies and approaches for their application in the scenarios where limited resources are available.

Examples of classification accuracies are provided in Figures 40 - 43.

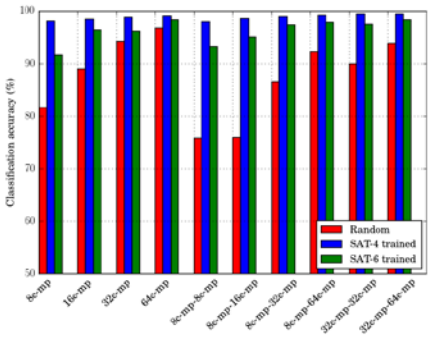


Figure 40 – Classification accuracies (%) on SAT-4 dataset obtained using features computed using convnets with one and two convolutional layers and varying number of convolutional units. Convnets are learned starting from randomly initialized weights.

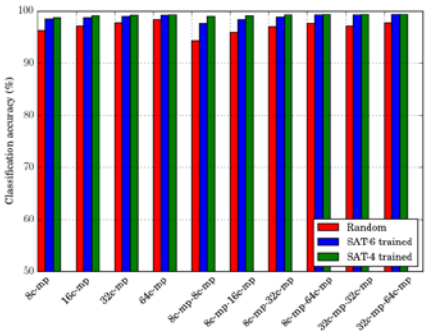


Figure 41 – Validation accuracies (%) on SAT-6 dataset obtained using features computed using convnets with one and two convolutional layers and varying number of convolutional units. Convnets are learned starting from randomly initialized weights.

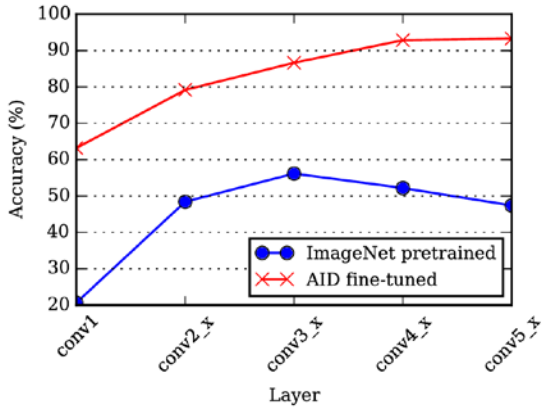


Figure 42 - Classification accuracies on AID dataset using ResNet for feature extraction and linear SVM for classification. Convnet is pretrained for object recognition on ImageNet dataset. Fine-tuning results in good classification accuracy when training set is small.

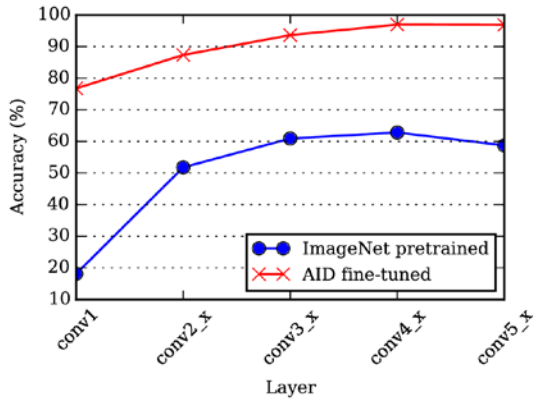


Figure 43 – Classification accuracies on UC Merced dataset using ResNet for feature extraction and linear SVM for classification. Convnet is pretrained for object recognition on ImageNet dataset and fine-tuned for HRRS image classification on AID dataset. No further fine-tuning on UC Merced dataset has been performed. One can see that the network can achieve excellent classification accuracy in spite of the differences in sensor type, image resolution and imaging conditions.

7.1.3.2 DATANMNH

Databases about National Museum of Natural History Collection
Digital Cultural Heritage – Area A

The images material related to some birds' species in the museum collection were uploaded to Avitohol cloud services.

The birds belong to the following taxonomic groups:

- Family Cracidae (Curassows), photo of 1 specimen (Crax mitu). An example is provided in Figure 44a.
- Family Trochilidae (Hummingbirds), photo of 42 specimens (Fhaetornis bourcieri; Melanotrochilus fuscus; Lampornis viridipallens; etc.). An example is provided in Figure 44b.

- Family Emberizidae (Buntings), photo of 29 specimens (Altapetes gutturalis; Emberiza calandra; Gubernatrix cristata; etc.). An example is provided in Figure 44c.
- Order Psittaciformes (Parrots), photo of 73 specimens (Trichoglossus haematodus; Conuropsis carolinensis; Amazona antomnalis; etc.). An example is provided in Figure 44d.
- Family Thraupidae (Tanagers), photo of 30 specimens (Agelaius xanthomus; Piranga rubra; Tangara guttata; etc.). An example is provided in Figure 44e.

The use of a supercomputer like Avitohol is very important for the Museum. This allows to manage the data in a very effective way. Having access to these resources is very important; however, improving the system usage and managing the data (upload, modify and download files) is also important.

There are 175 images of bird's species, uploaded in the Avitohol cloud service. These species are part of the National Museum of Natural History collection. Some of the images of the project are shown below.

a b c d e

Figure 44 – a. Family Cracidae (Curassows), (Crax mitu), b. Family Trochilidae (Hummingbirds), (Fhaetornis bourcieri), c. Family Emberizidae (Buntings), (Amblyrhamphus holosericeus), d. Order Psittaciformes (Parrots), (Callocephalon fibriatum), Family Thraupidae (Tanagers), (Chlorophanes spiza)

7.1.3.3 MC4CH

Management Cloud for Cultural Heritage
Digital Cultural Heritage – Area A

The aim of the project is to promote the understanding, appreciation, and management of heritage places, including:

- identification and inventory
- research and analysis
- monitoring and risk mapping
- determining needs and priorities for investigation, research, conservation and management
- planning for investigation, conservation, and management activities

- raising awareness and promoting understanding among the public, as well as governmental authorities and decision makers

Future work is planned to include getting to production phase, Automatic translation, importing 3D models and visualization, targeting possible users and creating user communities. MC4CH can be accessed through http://mc4ch.avitohol.acad.bg and a picture of the front page is provided in Figure 45.

Figure 45 – The front page of the Management Cloud for Cultural Heritage

7.1.3.4 BVL

Banatica Virtual Library
Digital Cultural Heritage – Area A

Main achievements of the project include:

- Sharing the meta-data in MARC 21 standard (XML, ISO, DAT formats) for more than 1000 books comprising the BVL collection; there is one file containing all entries
- Sharing 200 digitized old books in PDF format plus their covers as JPGs by making them available to DCH community via UVT's GridFTP (gsiftp://gridftp.viseem.hpc.uvt.ro) and DCH Repository (Clowder)
- Design and implementation of a distributed system for OCR Romanian prints
- Experiments on UVT's Cloud infrastructure (OpenStack)
- Experiments on AWS

Future work:

- Enhance OCR accuracy
- Create a vocabulary of old Romanian words extracted from the collection

This work unveils the old Romanian language and ease its study by linguists and philologists. The researchers can easily track the word dynamics and evolution. The Banat region, spanning across three countries, can track down its identity one hundred years back.

Central University Library “Eugen Todoran” and West University of Timisoara have identified few research directions in this area:

- Extend the OCR process to other digital collections (newspapers, brochures etc)
- Encourage and support researchers to use the digital repository in their work

An example of book covers provided through the VI-SEEM Clowder (http://dchrepo.vi-seem.eu) can be found in Figure 46.

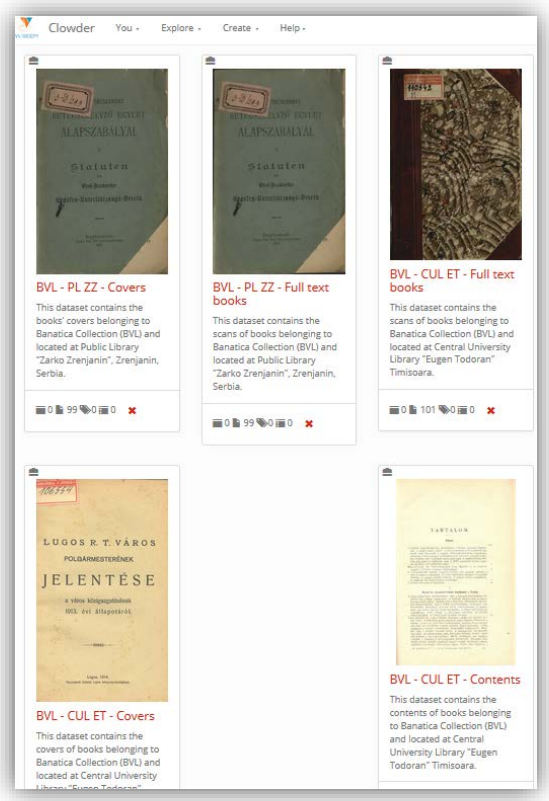


Figure 46 – BVL on Clowder. Some books' covers

7.2 List of publications from the 1st Open Call applications

1. Teimuraz Davitashvili, Nato Kutaladze, Ramaz Kvatadze, Georgi Mikuchadze, Zurab Modebadze, Inga Samkharadze, "*Precipitations Prediction by Different Physics of WRF Model*", International Journal of Environmental Science, Vol.1, 2016, pp.294-299, http://iaras.org/iaras/journals/ijes http://www.iaras.org/iaras/journals/caijes/precipitations-prediction-by-different-physics-of-wrf-model
2. Teimuraz Davitashvili, Nato Kutaladze, Ramaz Kvatadze, Georgi Mikuchadze, Zurab Modebadze, Inga Samkharadze, "*Showers Prediction by WRF Model above Complex Terrain*", Proceedings of the 39th International Convection MIPRO 2016/DC VIS, 2016, Opatija, Croatia, pp.236-241
3. Teimurazi Davitashvili, Nato Kutaladze, Ramaz Kvatadze and George Mikuchadze (2018) Scalable Computing: Practice and Experience, Volume 19, Number 2, pp. 199–208. http://www.scpe.org
4. Alvanos, M. and Christoudias, T.: "*GPU-accelerated atmospheric chemical kinetics in the ECHAM/MESSy (EMAC) Earth system model (version 2.52)*", Geosci. Model Dev., 10, 3679-3693, https://doi.org/10.5194/gmd-10-3679-2017, 2017
5. Alvanos, M. & Christoudias, T., (2017). "*MEDINA: MECCA Development in Accelerators – KPP Fortran to CUDA source-to-source Pre-processor"*. Journal of Open Research Software. 5(1), p.13. DOI: http://doi.org/10.5334/jors.158, 2017
6. Gadzhev, G., Georgieva, I., Ganev, K., Ivanov, V., Miloshev, N., Chervenkov, H., Syrakov, D. "*Climate applications in a virtual research environment platform*". Scalable Computing, Volume 19, Issue 2, 2018, Pages 107-118, SJR:0.180
7. Hristo Chervenkov, Vladimir Ivanov, Georgi Gadzhev, Kostadin Ganev. "*Sensitivity study of Different RegCM4.4 model set-ups – recent results from the TVRegCM experiment*". CYBERNETICS AND INFORMATION TECHNOLOGIES, Volume 17, No 5, 2017, pp. 17-26, SJR:0.203
8. Georgi Gadzhev, Vladimir Ivanov, Kostadin Ganev, and Hristo Chervenkov, "*TVRegCM Numerical Simulations - Preliminary Results*", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Volume 10665 LNCS, 2018, pp. 266-274. (SJR: 0.339)
9. D. Vudragovic, L. Ilic, P. Jovanovic, S. Nickovic, A. Bogojevic, and A. Balaz: "*VI-SEEM DREAMCLIMATE Service*"; Scal. Comput. Pract. Exp. 19, 215 (2018). DOI: 10.12694/scpe.v19i2.1396
10. Ivelina Georgieva, Georgi Gadzhev, Kostadin Ganev, Dimitris Melas, Tijian Wang. "*High Performance Computing Simulations of the Atmospheric Composition in Bulgaria and the City of Sofia*". CYBERNETICS AND INFORMATION TECHNOLOGIES, Volume 17, No 5, 2017, pp. 37-48, SJR:0.203
11. Ivelina Georgieva, Georgi Gadzhev, Kostadin Ganev, and Nikolay Miloshev., "*Computer Simulations of Atmospheric Composition in Urban Areas; Some Results for the City of Sofia*", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, Volume 10665, LNCS, pp. 474-482. (SJR: 0.339).

12. Giannaros TM, Kotroni V, Lagouvardos K, Matzarakis A (2018) "*Climatology and trends of the Euro-Mediterranean thermal bioclimate*", International Journal of Climatology, https://doi.org/10.1002/joc.5501.
13. Katragkou E., I. Gkotovou, Kartsios S., Pavlidis V., Tsigaridis K.,Trail M., Nazarenko L., Karacostas T., "*AUTH regional climate model contributions to EURO-CORDEX, Perspectives on Atmospheric Science*", Eds. Karacostas T. S., Bais A.F., Nastos P.T., p741-746, Springer International Publishing, Switzerland 2017.
14. Georgiou, G. K., Christoudias, T., Proestos, Y., Kushta, J., Hadjinicolaou, P., and Lelieveld, J.: "*Air quality modelling in the summer over the eastern Mediterranean using WRF-Chem: chemistry and aerosol mechanism intercomparison*", Atmos. Chem. Phys., 18, 1555-1571, https://doi.org/10.5194/acp-18-1555-2018, 2018.
15. Gkikas A, Giannaros TM, Kotroni V, Lagouvardos K, "*Assessing the radiative impacts of an extreme desert dust outbreak and the potential improvements on short-term weather forecasts: the case of February 2015*". Submitted for publication to Atmospheric Research.
16. V. Risojević, "*Analysis of learned features for remote sensing image classification*", In Proc. 13th Symposium on Neural Networks and Applications (NEUREL), pp.1-6. IEEE, 2016. (published).
17. R. Pilipović and V. Risojević, "*Evaluation of Convnets for Large-Scale Scene Classification From High-Resolution Remote Sensing Images*", In Proc. 17th International Conference on Smart Technologies, IEEE EUROCON 2017, pp. 932-937. IEEE, 2017. (published).
18. V. Stojnić and V. Risojević, "*Evaluation of Split-Brain Autoencoders for High-Resolution Remote Sensing Scene Classification*", In 60th International Symposium ELMAR-2018, IEEE, 2018. (accepted),
19. V. Stojnić and V. Risojević, "*Feature Learning for Remote Sensing Image Classification by Filling Missing Spectral Information*" (in preparation).
20. Miljan Bigovic, Luka Filipovic, Zarko Zacevic, Bozo Krstajic, "*Modeling and molecular dynamics simulations study of enol-carbonates and their derivatives*", Scalable Computing> practice and Experience, 2018, Vol. 19, No 2, 169-178.
21. Miljan Bigović, Žarko Zečević, Luka Filipović, Božo Krstajić, "*Verification of the three-dimensional structure of synthesized molecule by molecular dynamic simulations*", IEEE Eurocon 2017-17th International Conference of Smart Technologies, Ohrid, 6-8. July 2017, Book of abstracts 944-948.
22. Miljan Bigović, Žarko Zečević, Luka Filipović, Božo Krstajić, "*Komparacija računarskog modela molekula sa eksperimentalnim rezultatima*", XXII Međunarodni stručni skup ,Informacione tehnologije - sadašnjost i budućnost,, Žabljak, 27.02.-04.03.2017, Zbornik radova 189-192. ISBN 978-86-85775-20-8.
23. Bojana Koteska, Maja Simonoska Crcarevska, Marija Glavas Dodov, Jasmina Tonic Ribarska, and Ljupco Pejov. "*Semiempirical Atom-centered Density Matrix Propagation Approach to Temperature-dependent Vibrational Spectroscopy of Irinotecan*". In: Scalable Computing: Practice and Experience 19.2 (2018), pp. 149–159.
24. Bojana Koteska, Anastas Mishev, Marija Glavas Dodov, Maja Simonoska Crcarevska, Jasmina Tonic Ribarska, Vesna Petrovska Jovanovska, Monika Stojanovska, and Ljupco

Pejov. "*Modeling the solid-state vibrational spectroscopic properties of morphine-based formulations with hybrid meta density functional theory*". In: IEEE EUROCON 2017-17th International Conference on Smart Technologies. IEEE. 2017, pp. 938–943.

25. Bojana Koteska, Anastas Mishev, Ljupco Pejov, Maja Simonoska Crcarevska, Jasmina Tonic Ribarska, and Marija Glavas Dodov. "*Computational Vibrational Spectroscopy of Hydrophilic Drug Irinotecan*". In: Proceedings of the Eighth International Conference on Advances in System Simulation – SIMUL, 2016, pp. 11–16.
26. V. Petrovska Jovanovska, Lj. Pejov, A. Petrovska, S. Ugarkovic, M. Simonoska Crcarevska, M. Glavas Dodov. "*Physicochemical characterization and in vitro evaluation of modified release matrix tablets: The role of different grades of hydrohypropylmethyl cellulose*". Macedonian Pharmaceutical Bulletin, 63 (2018), in press.
27. B. Koteska, A. Mishev, Lj. Pejov. "*Pharmaceutical software quality assurance system*". In: Z. Budimac, Z. Horvath, T. Kozsik (eds.): Proceedings of the 5th Workshop of Software Quality Analysis, Monitoring, Improvement, and Applications (SQAMIA 2016), Budapest, Hungary, 29.-31.08.2016, vol. 1375, pp. 27-34. Also published online by CEUR Workshop Proceedings (CEURWS.org, ISSN 1613-0073).
28. Luka Ilic: "*VI-SEEM DREAMCLIMATE Service; e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean*", Sofia, Bulgaria, 15-16 May 2018.
29. Luka Ilic: "*DREAMCLIMATE application usage; Regional Climate training by the VI-SEEM project*", Belgrade, Serbia, 11-13 Oct 2017
30. N. Ivanova, A. Ivanova, "*Testing the limits of model membrane simulations-bilayer composition and pressure scaling*", J Comput Chem. (2018), 39(8):387-396
31. St. Iliev, G. Gocheva, N. Ivanova, B. Atanasova, J. Petrova, G. Madjarova, A. Ivanova, "*Identification and Computational Characterization of Isomers with cis and trans Amide Bonds in Folate and Analogues*"; Submitted.
32. J. Petrova, G. Gocheva, N. Ivanova, St. Iliev, B. Atanasova, G. Madjarova, A. Ivanova, "*Molecular Simulation of the Structure of Folate and Antifolates in Physiological Conditions*"; In Preparation.
33. G. Gocheva, A. Garcia Luri, St. Iliev, N. Ivanova, J. Petrova, B. Atanasova, A. Ivanova, G. Madjarova, "*Tautomerism in Folate: combined quantum and classical simulation*"; In Preparation.
34. N. Ivanova, A. Ivanova, "*Influence of the type of periodic boundary conditions on the transport of drug-peptide complex across model cell membranes*"; In Preparation.
35. N. Ivanova, A. Ivanova, "*Energy barriers for the transfer of a drug-peptide complex through lipid bilayers obtained from umbrella sampling simulations*"; In Preparation.
36. N. Ivanova, A. Ivanova, "*Elucidation of the penetration mechanism of a drug-CPP complex through a model erythrocyte membrane by atomistic molecular dynamics simulations*"; In Preparation
37. Athanasiou, C., S. Vasilakaki, D. Dellis, Z. Cournia. "*Using physics-based pose predictions and free energy perturbation calculations to predict binding poses and relative binding affinities for FXR ligands in the D3R Grand Challenge 2*". J. Comput.-Aided Mol. Des., 32(1): p. 21-44. Jan 2018.

38. Daniel Pop, Bogdan Irimie and Dana Petcu, “*Distributed Optical Character Recognition for Old Romanian Prints*”, Proc of the 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, IEEE CPS, Timisoara, September 2017.
39. G. Gadzhev and K. Ganev. Vertical structure of atmospheric composition fields over Bulgaria, International Conference on "*Numerical Methods for Scientific Computations and Advanced Applications*" (NMSCAA’18), Hisarya. Bulgaria, 27 – 31 May 2018, pp. 38-41.
40. I. Georgieva and N. Miloshev, “*Computer Simulations of PM Concentrations Climate for Bulgaria*”, International Conference on "Numerical Methods for Scientific Computations and Advanced Applications" (NMSCAA’18), Hisarya. Bulgaria, 27 – 31 May 2018, pp. 46-49.
41. Gadzhev, G., Georgieva, I., Ganev, K., Miloshev, N., “*CONTRIBUTION OF DIFFERENT EMISSION SOURCES TO THE ATMOSPHERIC COMPOSITION FORMATION IN CITY OF SOFIA*”, HARMO 2018 - 18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Proceedings, (2017), pp. 236-240.
42. Georgieva, I., Ivanov, I., “*IMPACT OF THE AIR POLLUTION ON THE QUALITY OF LIFE AND HEALTH RISKS IN BULGARIA*”, HARMO 2018 - 18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Proceedings, (2017), pp. 647-652.
43. Ivanov, V. and Georgieva, I., (2017) “Air quality index evaluations for Sofia city”, 17th IEEE International Conference on Smart Technologies, EUROCON 2017 - Conference Proceedings.

7.3 Measurable data from the 1st Open Call applications

7.3.1 Usage of Resources in the 1st Open Call

After a detailed technical and a brief scientific evaluation, project's consortium accepted 21 applications (projects) in total in the 1st call for production use of VI-SEEM resources and services. The majority of the projects (14 out of 20) requested access to HPC, six of them requested access to Cloud resources, one of them (VINE application from Georgia) needed Grid infrastructure, and one project (AUTH_WRF371M_EUR0.44 application from Greece) was granted access to the VI-SEEM repository service only. During the access phase, DIOPTRA application developers changed the scope of their project and decided to withdraw from the use of awarded resources. Therefore, 20 applications in total have used the allocated infrastructure resources.

Figure 47 illustrates the utilization of allocated HPC CPU resources. Blue bars represent amounts of allocated CPU resources, while red bars show the corresponding usage per application. Applications accepted in the first call significantly utilized the allocated resources, up to 98% of the allocation. In absolute numbers, the projects consumed 10 out of 10.2 million of the assigned CPU-core hours. During this call, the applicants overestimated their needs for GPU resources, which resulted in a consumption of just about 52% (650 thousands of GPU-card hours) of allocated resources. On the other hand, the requests for the Xeon Phi-card hours have turned out to be underestimated. With the consent of our colleagues from IICT-BAS (Avitohol cluster), the initially allocated resources were significantly increased (6 times, from 1 to 6.3 million of Xeon Phi-core hours) and utilized by the ACIQLife project. In the case of CH-CBIR application, although initially only GPU resources were allocated, additional CPU resources became necessary to the project. In this and similar circumstances, the project's consortium approved additional, reasonable amounts of resources.

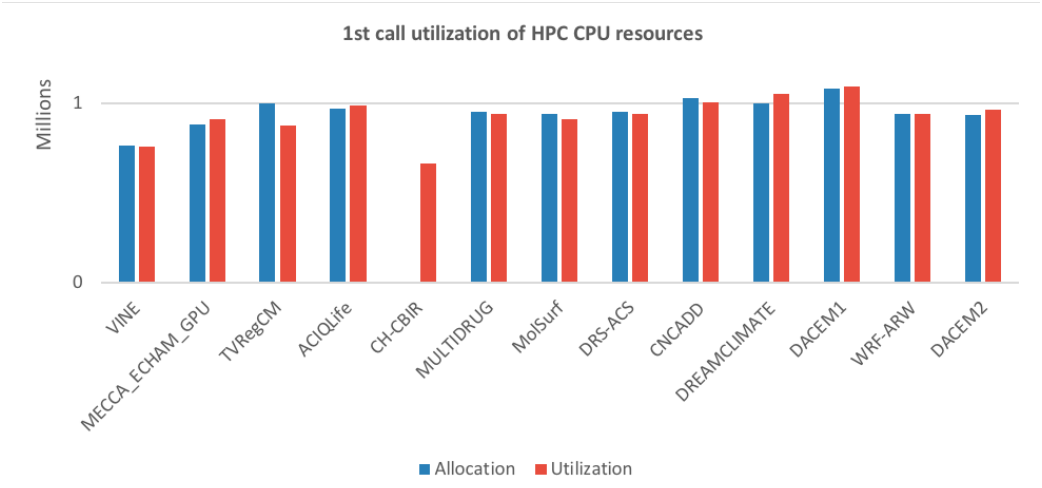


Figure 47 – Utilization of allocated HPC CPU resources.

The total utilization of allocated Cloud resources stands at approximately 95%, and the first call projects consumed 78 out of 82 allocated VM-cores. Figure 48 illustrates the numbers assigned (in blue) and used (in red) VM-cores per project.

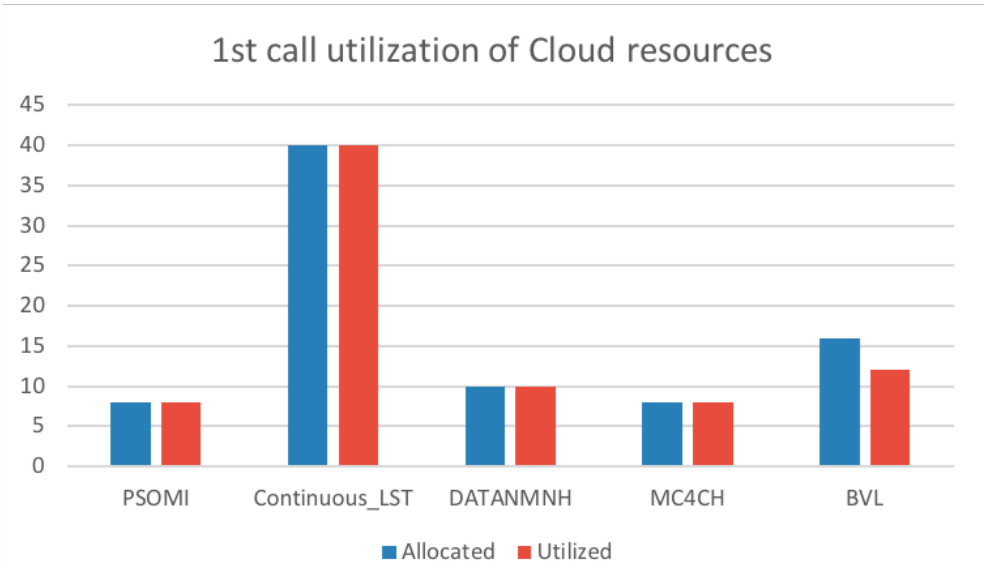


Figure 48 – Utilization of allocated Cloud resources.

Only one project from the first call for production use of VI-SEEM resource and services used Grid infrastructure (VINE from Georgia). Through this interface, the project consumed around 36,000 of CPU-core hours.

The total storage space consumption by the projects from the first call on HPC storage systems amounts to 67.2 TB. In addition to this, approximately 10 TB of storage space is occupied at the Simple storage service, 3TB of storage space at the Archival service, and in total 3 TB of storage space at the application- and data-specific services, which include: the Data repository service, the Live Access Server, and the Clowder.

7.3.2 PhD / MSc Thesis and Talks / Posters resulted from the 1st Open Call

Actions	Number
PhD/MSc thesis (taking advantage of / collaborating in the developed project)	9
Talks given/Posters presented	59

7.3.2.1 PhD/MSc Thesis

The following PhD thesis taking advantage of / collaborating in the developed project are the following:

1. Ivelina Georgieva – "*Local transport processes and chemical transformation in the atmosphere*", PhD.
2. George K. Georgiou: "*Air Quality Modelling over Cyprus*", PhD.
3. Nikoleta Ivanova, "*Study of the penetration of a drug-peptide complex trough model cell membranes*", PhD.
4. Vesna Petrovska Jovanovska, "*Development and Formulation optimisation of Modified release dosage form using quality by design – QdD approach*", PhD
5. Bojana Koteska, "*Framework for developing scientific applications*", PhD
6. Inga Samkharadze – "*Environmental pollution and local atmospheric processes mathematical modelling*", PhD.
7. Gkotovou Ioanna – "*Impact of boundary forcing on regional climate simulations and evaluation with satellite data*", MSc.
8. Verce Manevska – "*Quantum mechanical modeling of intra- and intermolecular motions in formic acid and its noncovalently bonded dimers with benzene*", MSc
9. Bogdan Irimie has been working on his PhD thesis through the period.

7.3.2.2 Talks/Posters

The following Talks/Posters have been presented:

1. Teimuraz Davitashvili, Nato Kutaladze, Ramaz Kvatadze , Giorgi Mikuchadze, Zurab Modebadze, "*Heavy showers prediction above the complex terrain based on WRF modelling*", VII International Joint Conference of Georgian Mathematical Union & Georgian Mechanical Union "Continuum Mechanics and Related Problems of Analysis" Dedicated to 125-th birthday anniversary of academician N. Muskhelishvili, 5-9 September, 2016, Batumi, Georgia, http://www.gmu.ge/Batumi2016/Files/poster_Eng.pdf
2. Teimuraz Davitashvili, "*Showers prediction by WRF model above complex terrain*", The 1st Eastern Partnership E-Infrastructure Conference (EPEC 2016), University of Georgia, October 6-7, 2016, Tbilisi, Georgia.
3. Teimuraz Davitashvili, "*Mathematical Simulation of Possible Detrimental Event Development above the Territory of Georgia*", International Conference dedicated to Anniversary of Birthday of G.A.Aliyev, Baku, May 22-25, 2016.
4. Teimuraz Davitashvili, "Precipitations Prediction by Different Physics of WRF Model", 9thInternational Conference on Energy Planning, Energy Saving, Environmental Education (EPESE '16) , Rome, Italy, October 21-23, 2016, http://www.wseas.org/cms.action?id=12981
5. Teimuraz Davitashvili, "*Investigation of Some Unfavorable Atmosphere Processes Development on the Territory of Georgia by WRF Modeling*", 9th International Conference on Energy Planning, Energy Saving, Environmental Education (EPESE '16), Rome, Italy, October 21-23, 2016, http://www.wseas.org/cms.action?id=12981
6. Ramaz Kvatadze. "e-Infrastructure for Research and Education in Georgia", VIII Anual International Conference of the Georgian Mathematical Union, 4-8 September,2017, Batumi, Georgia.
7. Nato Kutaladze and Nino Shareidze, "*Regional climate prediction system for south Caucasus region*", VIII Anual International Conference of the Georgian Mathematical Union, 4-8 September,2017, Batumi, Georgia.
8. George Mikuchadze, "*Simulation of Air Pollutant Distribution over the Caucasus on the bases of WRF-Chem model*", VIII Anual International Conference of the Georgian Mathematical Union, 4-8 September,2017, Batumi, Georgia.
9. Zurab Modebadze, "*Software of distributed computing network monitoring and analysis*", VIII Anual International Conference of the Georgian Mathematical Union, 4-8 September,2017, Batumi, Georgia.
10. Teimuraz Davitashvili, "*Simulation and analysis of some non-ordinary atmosphere processes by WRF model based on the GRID Technologies*", VIII Anual International Conference of the Georgian Mathematical Union, 4-8 September,2017, Batumi, Georgia.
11. Inga Samkharadze and Teimuraz Davitashvili, "*Convective clouds prediction based on ARL aerological diagrams and radar observations data analysis*", VIII Anual

International Conference of the Georgian Mathematical Union, 4-8 September,2017, Batumi, Georgia.

12. Teimuraz Davitashvili, Nato Jiadze, Zurab Modebadze, "*Showers Simulation Study over Caucasus Region by WRF Model Based on Grid Computing*", The International Conference "Mathematical Modeling and Computational Physics, 2017" (MMCP2017) 3 - 7 July 2017, Dubna, Russia.
13. Teimuraz Davitashvili, "*Effect of dust aerosols in forming the regional climate of Georgia*", International Conference on EU TEMPUS project: "Modernization of Mathematics curricula for Engineering and Natural Sciences studies in South Caucasian Universities by introducing modern educational technologies" 20 February, 2017, Tbilisi, Georgia.
14. Luka Ilic, "*VI-SEEM DREAMCLIMATE Service; e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean*", Sofia, Bulgaria, 15-16 May 2018.
15. Luka Ilic, "*DREAMCLIMATE application usage; Regional Climate training by the VI-SEEM project*", Belgrade, Serbia, 11-13 Oct 2017.
16. Michail Alvanos and Theodoros Christoudias, "*Accelerated chemical kinetics in the EMAC chemistry-climate model*", International Conference on High Performance Computing & Simulation (HPCS), 2016.
17. Theodoros Christoudias, "*EMAC Accelerated Atmospheric Climate Kinetics in Earth System Modelling*", "e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean" Conference, 2018.
18. J Kushta, Y Proestos, G Georgiou, T Christoudias, J Lelieveld, "*Application of the WRF-Chem model for the simulation of air quality over Cyprus*" EGU General Assembly Conference Abstracts 19, 12333,
19. T Christoudias, Y Proestos, J Kushta, P Hadjinicolaou, J Lelieveld, "*Air quality modelling over the Eastern Mediterranean using the WRF/Chem model: Comparison of gas – phase chemistry and aerosol mechanisms*", GK Georgiou, EGU General Assembly Conference Abstracts 19, 7894.
20. Presentation "*Structure of dense adsorption layers of escin at the air-water interface studied by molecular dynamics simulations*" in the Workshop "Two Years Avitohol: Advanced HPC applications" 29-31 October, 2017, Panagyurishte, Bulgaria.
21. Presentation "*Molecular dynamics study of the kinetics of adsorption of LAS molecules*" in the Workshop "Two Years Avitohol: Advanced HPC applications" 29-31 October, 2017, Panagyurishte, Bulgaria.
22. Bojana Koteska, Anastas Mishev, and Ljupco Pejov, "*Computational approach towards vibrational spectroscopic detection of molecular species relevant to atmospheric chemistry and climate science: The formic acid rotamers*". IEEE EUROCON 2017-17th International Conference on Smart Technologies.
23. Ljupco Pejov. "*Efficient exact computations of anharmonic X-Y stretching frequencies for molecular species relevant to atmospheric chemistry, climate science and drug delivery*". VI-SEEM regional climate training event, 11-13 October 2017, Institute of Physics, Belgrade.
24. Bojana Koteska, Verce Manevska, Jovica Todorov, Anastas Mishev, and Ljupco Pejov. "*Dynamic versus Static Approach to Theoretical Anharmonic Vibrational Spectroscopy*

of Molecular Species Relevant to Atmospheric Chemistry: A Case Study of Formic Acid". E-infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean, 15-16 May 2018, Sofia.

25. Georgi Gadzhev, Ivelina Georgieva, Kostadin Ganev, Vladimir Ivanov, Nikolay Miloshev, Hristo Chervenkov and Dimiter Syrakov, "*Climate applications in a virtual research environment platform*", International conference "e-Infrastructures for excellent science in Southeast Europe and Eastern Mediterranean" 15-16 May 2018 Sofia, Bulgaria.
26. Hristo Chervenkov, Vladimir Ivanov, Georgi Gadzhev, Kostadin Ganev, "*Sensitivity study of Different RegCM4.4 model set-ups – recent results from the TVRegCM experiment*", Workshop "Two Years Avitohol: Advanced HPC applications", 29-31 October 2017, Panagyurishte, Bulgaria.
27. Gadzhev, G., Georgieva, I., Ganev, K., Ivanov, V., Miloshev, N., Chervenkov, H., Syrakov, D. "*Climate applications in a virtual research environment platform*", 11th International Conference on Large-Scale Scientific Computations June 5 - 9, 2017, Sozopol, Bulgaria (LSSC'17).
28. Giannaros TM (2017) "*On the impact of dust on Mediterranean cyclones: the Medicane case of December 2005*". 18th Cyclone Workshop, 1 – 6 October 2017, Saint Adele, Quebec, Canada.
29. Giannaros TM (2017) "On the impact of dust on Mediterranean cyclones: the Medicane case of December 2005". 10th HyMeX Workshop, 4 – 7 July 2017, Barcelona, Spain.
30. Hristo Chervenkov, Vladimir Ivanov, Georgi Gadzhev, Kostadin Ganev, "*Sensitivity Study of Different RegCM4.4 Model Configurations – Consolidated Results from the TVRegCM Experiment*", International conference "e-Infrastructures for excellent science in Southeast Europe and Eastern Mediterranean" 15-16 May 2018 Sofia, Bulgaria.
31. Georgi Gadzhev, Ivelina Georgieva, Kostadin Ganev, Vladimir Ivanov, Nikolay Miloshev, Hristo Chervenkov and Dimiter Syrakov, "*Climate applications in a virtual research environment platform, International conference*", "e-Infrastructures for excellent science in Southeast Europe and Eastern Mediterranean" 15-16 May 2018 Sofia, Bulgaria.
32. G. Gadzhev and K. Ganev, "*Vertical structure of atmospheric composition fields over Bulgaria*", International Conference on "Numerical Methods for Scientific Computations and Advanced Applications" (NMSCAA'18), Hisarya. Bulgaria, 27 – 31 May 2018.
33. Georgieva and N. Miloshev, "*Computer Simulations of PM Concentrations Climate for Bulgaria*", International Conference on "Numerical Methods for Scientific Computations and Advanced Applications" (NMSCAA'18), Hisarya. Bulgaria, 27 – 31 May 2018.
34. Ivelina Georgieva, Georgi Gadzhev, Kostadin Ganev, Dimitris Melas, Tijian Wang, "*High Performance Computing Simulations of the Atmospheric Composition in Bulgaria and the City of Sofia*", Workshop "Two Years Avitohol: Advanced HPC applications", 29-31 October 2017, Panagyurishte, Bulgaria.
35. Gadzhev, G., Georgieva, I., Ganev, K., Miloshev, N., "*CONTRIBUTION OF DIFFERENT EMISSION SOURCES TO THE ATMOSPHERIC COMPOSITION FORMATION IN CITY OF SOFIA*", HARMO 2018 - 18th International Conference on Harmonisation within

Atmospheric Dispersion Modelling for Regulatory Purposes, 9-12 October 2017, Bologna, Italy.

36. Georgieva, I., Ivanov, I., "IMPACT OF THE AIR POLLUTION ON THE QUALITY OF LIFE AND HEALTH RISKS IN BULGARIA", HARMO 2018 - 18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 9-12 October 2017, Bologna, Italy.
37. Ivanov, V. and Georgieva, I. "Air quality index evaluations for Sofia city", 17th IEEE International Conference on Smart Technologies, EUROCON 2017.
38. Ivelina Georgieva, Georgi Gadzhev, Kostadin Ganev, and Nikolay Miloshev., "*Computer Simulations of Atmospheric Composition in Urban Areas: Some Results for the City of Sofia*", 11th International Conference on Large-Scale Scientific Computations June 5 - 9, 2017, Sozopol, Bulgaria (LSSC'17).
39. Giannaros TM (2018) "*The Euro-Mediterranean thermal bioclimate: Climatology and trends*", 11th HyMeX Workshop, 29 May – 2 June 2018, Lecce, Italy.
40. Giannaros TM (2017) "*The WRF-Chem model*", VI-SEEM REG CL: Regional training event, 11 October 2017, Belgrade, Serbia.
41. Miljan Bigovic, "*Modeling and molecular dynamics simulations study of enol-carbonates and their derivatives*", Sofia, Bulgaria, 14-17. 05. 2018.
42. G. Gocheva, S. Iliev, B. Atanasova, N. Ivanova, J. Petrova, A. Garcia Luri, G. Madjarova, A. Ivanova, "*Computational investigation of the molecular structure of folic acid and antifolates in aqueous solution*" (poster), Computational Advances in Drug Discovery, 05-08. 09. 2017, Lausanne, Switzerland.
43. G. Gocheva, S. Iliev, B. Atanasova, N. Ivanova, J. Petrova, A. Garcia Luri, G. Madjarova, A. Ivanova, "*Structural characterization of folic acid and antifolates in aqueous solution*", International Workshop "Advanced Materials", 10-13. 09. 2017, Pomorie, Bulgaria.
44. G. Gocheva, S. Iliev, B. Atanasova, N. Ivanova, J. Petrova, G. Madjarova, A. Ivanova, "*Folic acid and anifolates in aqueous solution – a molecular dynamics study*" (talk), Workshop "Two Years Avitohol: Advanced HPC applications", 29-31. 09. 2017, Panagyurishte, Bulgaria.
45. N. Ivanova, A. Ivanova, "*Transfer of a drug-peptide complex through model cell membranes*" (talk), Workshop "Two Years Avitohol: Advanced HPC applications", 29-31. 09. 2017, Panagyurishte, Bulgaria.
46. S. Iliev, G. Gocheva, B. Atanasova, N. Ivanova, J. Petrova, A. Garcia Luri, G. Madjarova, A. Ivanova, "*Quantum chemical characterization of folate and antifolates*" (talk in Bulgarian), IV Science conference for students, PhD students and young scientists "The challenges in chemistry", 10-11. 10. 2017, Plovdiv, Bulgaria.
47. N. Ivanova, A. Ivanova, "*Exploring the possibilities for penetration of a drug-peptide complex into model membrane systems*" (talk), HUMBOLDT KOLLEG: Humboldtians and scientific progress in the Central and East European countries, 16-18. 11. 2017, Sofia, Bulgaria.
48. B. Atanasova, S. Iliev, G. Gocheva, N. Ivanova, J. Petrova, G. Madjarova, A. Ivanova, "*Folic acid and antifolates – from molecular characteristics to biological activity*"

(poster), Scientific session of the Faculty of chemistry and pharmacy of Sofia University "St. Kliment Ohridski", 24. 11. 2017, Sofia, Bulgaria.

49. J. Petrova, G. Gocheva, S. Iliev, B. Atanasova, N. Ivanova, G. Madjarova, A. Ivanova, "*Folic Acid and Antifolates as Targeting Ligands of the α-folate Receptor – a Theoretical Insight*" (poster), 19th International Workshop on Nanoscience and Nanotechnology NANO 2017, 24-25. 11. 2017, Sofia, Bulgaria.
50. A. Ivanova, G. Gocheva, S. Iliev, B. Atanasova, N. Ivanova, J. Petrova, G. Madjarova, "First modelling steps of folate and derivatives as prospective vector ligands" (talk), Seminar of the Department of physical chemistry of Faculty of chemistry and pharmacy of Sofia University "St. Kliment Ohridski", 29. 01. 2018, Sofia, Bulgaria.
51. G. Gocheva, S. Iliev, B. Atanasova, N. Ivanova, J. Petrova, G. Madjarova, A. Ivanova, "*Structural differences between cis and trans folic acid and antifolates in physiological environment – a theoretical study*" (talk), e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean, 15-16. 05. 2018, Sofia, Bulgaria.
52. Bogdan Irimie, "*BVL: Distributed Optical Character Recognition for Old Romanian Prints*", International conference "e-Infrastructures for excellent science in Southeast Europe and Eastern Mediterranean", Sofia, May 2018.
53. Bojana Koteska, Anastas Mishev, Marija Glavas Dodov, Maja Simonoska Crcarevska, Jasmina Tonic Ribarska, Vesna Petrovska Jovanovska, Monika Stojanovska, and Ljupco Pejov, "*Modeling the solid-state vibrational spectroscopic properties of morphine-based formulations with hybrid meta density functional theory*". IEEE EUROCON 2017-17th International Conference on Smart Technologies.
54. Bojana Koteska, Anastas Mishev, Ljupco Pejov, Maja Simonoska Crcarevska, Jasmina Tonic Ribarska, and Marija Glavas Dodov. "*Computational Vibrational Spectroscopy of Hydrophilic Drug Irinotecan*". Eighth International Conference on Advances in System Simulation – SIMUL, 2016.
55. B. Koteska, A. Mishev, Lj. Pejov. "*Pharmaceutical software quality assurance system*". 5th Workshop of Software Quality Analysis, Monitoring, Improvement, and Applications (SQAMIA 2016), Budapest, Hungary, 29.-31.08.2016.
56. Bojana Koteska, Maja Simonoska Crcarevska, Marija Glavas Dodov, Jasmina Tonic Ribarska, and Ljupco Pejov. "*Semiempirical Atom-centered Density Matrix Propagation Approach to Temperature-dependent Vibrational Spectroscopy of Irinotecan*". E-infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean, 15-16 May 2018, Sofia.
57. Georgieva, G. Gadzhev, K. Ganev, N. Miloshev, "Computer simulations of PM pollution in urban areas - some results for Sofia city", European Geosciences Union General Assembly 2018 Vienna, Austria, 8–13 April 2018 (EGU 18).
58. D. Agre, G. Agre, I. Georgiev, V. Pirinski, "*Inventory and Management Cloud for Bulgarian Cultural Heritage*", International conference "e-Infrastructures for excellent science in Southeast Europe and the Eastern Mediterranean", 15-16 May, 2018, Sofia, Bulgaria.
59. Georgiev, V. Pirinski, "*Inventory and Management Cloud for Cultural Heritage*", VI-SEEM Digital cultural heritage training event at Bibliotheca Alexandrina, 6-8 February, 2017, Alexandria, Egypt.

8 Conclusion

This document provides the report on the 1st and 2nd calls for proposals for projects accessing VI-SEEM resources and services, the results of the review process of the applications, the applications which have been given access to computational resources, storage resources and VI-SEEM services as well as the efforts of supporting the chosen applications. WP6 oversees capacity building and, therefore, supports open access calls for VRE partners using existing data and codes but also providing new ones, enriching further the services of the VRE. In total four open calls will be organised and this script focusses only on the first two calls. The 3rd call for proposals for projects accessing VI-SEEM resources and services has now closed and successful projects have been granted resources.

The 1st and 2nd open calls has been designed and implemented according to the outcome of deliverable D6.1 “Framework for VRE resource and service provision”. The call was addressed to scientists and researchers that work in academic and research institutions in the region of South Eastern Europe and the Eastern Mediterranean and focuses on research topics in the specific fields of Life Sciences, Climate research, and Digital Cultural Heritage.

The first call enabled researchers from SEEM countries to access VI-SEEM resources and services. In total 22 applications were received, underwent technical review and where all the requirements were clarified they were sent for the scientific review. 21 applications appeared to match the evaluation criteria, and were supported. However, one application withdrew from the usage of resources. Hence, 20 applications made use of VI-SEEM resources and services namely, 11 in Climate Research, 5 in Life Sciences and 4 in Digital Cultural Heritage.

The second call enabled researchers from SEEM countries to access VI-SEEM resources and services. In total 19 applications were received, underwent technical review and where all the requirements were clarified they were sent for the scientific review. 17 applications appeared to match the evaluation criteria, and were supported. Hence, 17 applications got access to VI-SEEM resources and services namely, 4 in Climate Research, 7 in Life Sciences and 6 in Digital Cultural Heritage.

In addition each SEEM country has established a Service Enabler the role of which is to guide the primary investigators of each project to access the allocated resources and services as well as to oversee the progression of the projects.

Final report will be available in D6.4.

