# LATE REQUEST FOR A SPECIAL PROJECT 2018–2020

**MEMBER STATE:** Greece, France ..................................................................

This form needs to be submitted via the relevant National Meteorological Service.

**Principal Investigator[1]:** Vassilios D. Vervatis (1), Pierre De Mey (2) ............................

**Affiliation:** (1) National Kapodistrian University of Athens (UoA)................
(2) Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)..........................................................

**Address:** (1) Ocean Physics And Modelling group (OPAM), Faculty of Physics, Division of Environmental Physics and Meteorology, Campus Building PHYS-V, 15784 Athens, Greece....................
(2) LEGOS/CNES 18, av. Edouard Belin, 31401 Toulouse cédex 9, France ..........................................................

**E-mail:** (1) vervatis@oc.phys.uoa.gr ................................................
(2) pierre.de-mey@legos.obs-mip.fr ....................................

**Other researchers:** Sarantis Sofianos (1), Nadia Ayoub (2), Marios Kailas (1)

**Project Title:** Stochastic Coastal/Regional Uncertainty Modelling (SCRUM) 2: consistency, reliability, probabilistic forecasting, and contribution to CMEMS ensemble data assimilation....................................

| | | |
|---|---|---|
| If this is a continuation of an existing project, please state the computer project account assigned previously. | **SPGRVERV** | |
| Starting year:<br>(A project can have a duration of up to 3 years, agreed at the beginning of the project.) | 2018 | |
| Would you accept support for 1 year only, if necessary? | YES ☒ | NO ☐ |

| **Computer resources required for the years:**<br>(To make changes to an existing project please submit an amended version of the original form.) | | **2018** | **2019** | **2020** |
|---|---|---|---|---|
| High Performance Computing Facility | (SBU) | 5.000.000 | 8.000.000 | 2.000.000 |
| Accumulated data storage (total archive volume)[2] | (GB) | 7.000 | 15.000 | 18.000 |

***An electronic copy of this form must be sent via e-mail to:*** *special_projects@ecmwf.int*

Electronic copy of the form sent on (please specify date): March 09, 2018 ..................................................

*Continue overleaf*

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide an annual progress report of the project's activities, etc.

[2] If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year.

# LATE REQUEST FOR A SPECIAL PROJECT 2018–2020

| | |
|---|---|
| **MEMBER STATE:** | Greece, France ………………………………………………………...…<br>This form needs to be submitted via the relevant National Meteorological Service. |
| **Principal Investigator[1]:** | Vassilios D. Vervatis (1), Pierre De Mey (2) ………….……...… |
| **Affiliation:** | (1) National Kapodistrian University of Athens (UoA)..…………<br>(2) Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS)..………………………………………………… |
| **Address:** | (1) Ocean Physics And Modelling group (OPAM), Faculty of Physics, Division of Environmental Physics and Meteorology, Campus Building PHYS-V, 15784 Athens, Greece…………………<br>(2) LEGOS/CNES 18, av. Edouard Belin, 31401 Toulouse cédex 9, France ………………………………………………………. |
| **E-mail:** | (1) vervatis@oc.phys.uoa.gr ……………………………………...<br>(2) pierre.de-mey@legos.obs-mip.fr ……………………………….. |
| **Other researchers:** | Sarantis Sofianos (1), Nadia Ayoub (2), Marios Kailas (1) |
| **Project Title:** | Stochastic Coastal/Regional Uncertainty Modelling (SCRUM) 2: consistency, reliability, probabilistic forecasting, and contribution to CMEMS ensemble data assimilation……………………………... |

| | | |
|---|---|---|
| If this is a continuation of an existing project, please state the computer project account assigned previously. | **SPGRVERV** | |
| Starting year:<br>(A project can have a duration of up to 3 years, agreed at the beginning of the project.) | 2018 | |
| Would you accept support for 1 year only, if necessary? | YES ☒ | NO ☐ |

| **Computer resources required for the years:**<br>(To make changes to an existing project please submit an amended version of the original form.) | | **2018** | **2019** | **2020** |
|---|---|---|---|---|
| High Performance Computing Facility | (SBU) | 5.000.000 | 8.000.000 | 2.000.000 |
| Accumulated data storage (total archive volume)[2] | (GB) | 7.000 | 15.000 | 18.000 |

***An electronic copy of this form must be sent via e-mail to*:** *special_projects@ecmwf.int*

Electronic copy of the form sent on (please specify date):

…………………………………………

*Continue overleaf*

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide an annual progress report of the project's activities, etc.

[2] If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year.

**Principal Investigator:** Vassilios D. Vervatis and Pierre De Mey ..............................

**Project Title:** Stochastic Coastal/Regional Uncertainty Modelling (SCRUM) 2: consistency, reliability, probabilistic forecasting, and contribution to CMEMS ensemble data assimilation ...............................

# Extended abstract

*All Special Project requests should provide an abstract/project description including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used.*

*Requests asking for 1,000,000 SBUs or more should be more detailed (3-5 pages).*

*Following submission by the relevant Member State the Special Project requests will be based on the following criteria: Relevance to ECMWF's objectives, scientific and technical quality, disciplinary relevance, and justification of the resources requested. Previous Special Project reports and the use of ECMWF software and data infrastructure will also be considered in the evaluation process.*

*All accepted project requests will be published on the ECMWF website.*

**Abstract**

The work proposed here builds upon, and expands, the previous ECMWF Special Project (account id SPGRVERV) and a joint project within CMEMS Service Evolution named SCRUM (https://www. mercator-ocean.fr/portfolio/scrum-2/). The new work, funded by CMEMS under the project name SCRUM2, is based on stochastic modelling of ocean physics and biogeochemistry, in the context of coastal/regional Ensemble DA forecasting systems, and includes methods suitable to assess the reliability of Ensembles in probabilistic assimilation systems. It aims at strengthening CMEMS in the areas of regional/coastal ocean uncertainty modelling, Ensemble consistency verification, Ensemble probabilistic forecasting, and Ensemble data assimilation. In a first step, we will revisit Ensemble generation and sensitivity analysis approaches by accounting for the age of errors and considering the oceanic response to ECMWF-EPS atmospheric Ensembles. In a second step, we will (1) perform and expand our Ensemble consistency analysis to correlated observational errors, and (2) develop methods to assess the reliability of Ensembles in probabilistic assimilation systems.

## 1. Introduction

Ocean forecasting systems, such as those used in CMEMS for service-related activities, merge information from dynamical and ecosystem models and observational networks through data assimilation in order to produce suitable, quality-controlled ocean state estimates and forecasts. In recent years, notably in the MyOcean line of European R&D projects, and other projects, considerable effort has been dedicated to reduce forecast errors based on model improvement, observational array design and evolution of data assimilation methods, within the context of ever increasing computing capacities.

The need for specific data assimilation methods and sustained observations in regional models and in coastal regions of larger-scale models has been identified early enough (see extensive reviews in De Mey, 2000, and Pinardi et al., 2017). First of all, there are major differences between deep-ocean and coastal ocean processes to be estimated. Continental shelf and slope seas differ from the open ocean in the presence of the coast, strong bathymetry gradients, inputs from rivers, and shallower water. Coastal-trapped waves propagate cyclonically around the ocean basin, on the gradient of potential vorticity caused by this depth change (Huthnance, Mysak and Wang, 1986). Flows in these shallower seas are forced by (inter alia) pressure and current fields from ocean-scale mass balances, circulation, tides and eddies, winds and air pressure variations, non-uniform density (due to solar heating, river inputs, precipitation-evaporation, latent and sensible heat fluxes). For all these, responses differ between the deep ocean and shallower shelf. Specific processes include:

slope current and associated mesoscale and sub-mesoscale shelf break exchanges, shelf dynamics, fronts, storm surges, tides, internal waves, surface waves, swell, upwellings, fluxes of nutrients, sediments and pollutants, estuarine processes, plumes, topographic controls on circulation. Also, the influence of coastal ocean processes is felt far beyond shelf break, thus interacting with open ocean dynamics and controlling the connectivity of remote ecosystems.

In the coastal parts of regional ocean models, many factors can complicate the assimilation of data compared to the deep-ocean: free-surface variations (tides, storm surges), anisotropy (offshore scales are generally shorter than alongshore scales), non-homogeneity, friction and mixing effects throughout the water column (driven in part by tides), shallow water and strong variations of bathymetry, freshwater input and river plumes, and the fast response to the atmosphere on the shelf. The characterization and specification of model error are critical in any assimilation scheme but extremely challenging in the coastal zone. The model errors are strongly dependent on time scale but any attempt at separation is confounded by strong nonlinearity in the dynamics that can couple variability at different frequencies. As a first step, one must characterize the forecast errors under various error regimes by methods which include realistic error dynamics such as stochastic modelling (e.g., Quattrocchi et al., 2015; Vervatis et al., 2016). Most studies point at the benefit of "advanced" assimilation methods with built-in error propagation (Kourafalou et al., 2015a,b), such as the Ensemble Kalman filter and other Ensemble methods.

## 2. Scientific Objectives

An Ensemble approach based on a high-resolution Bay of Biscay NEMO configuration, has been installed at ECMWF HPCF by the University of Athens and LEGOS/CNRS research teams. The currently ongoing SP, to end this year in 2018, aimed to advance the approach on the following topics:

- Consolidate the stochastic modelling work which has already been undertaken by using up-to-date perturbation and analysis techniques.
- Expand the stochastic modelling work to ecosystem-sensitive variables within a coupled ocean-biogeochemical forecasting system.
- Introduce and test two methodologies aimed at checking the consistency of Ensembles with respect to CMEMS Thematic Assembly Centres (TAC) data and arrays.
- Showcase the use of those Ensemble-modelled uncertainties in a pilot data assimilation exercise.
- Contribute guidance in support of upcoming decisions regarding the evolution of regional/coastal data assimilation schemes in CMEMS-e.g. suboptimality, analysis scheme, etc.

In this new proposed SP, we wish to continue and expand the work undertaken in a joint Copernicus marine research project (https://www.mercator-ocean.fr/portfolio/scrum-2/), focussing more precisely on specific topics:

- The sensitivity of simulations to uncertainties and generation of Ensembles with account for the age of errors and in response to ECMWF-EPS atmospheric Ensembles (Fig. 1).
- Ensemble reliability and forecast probability diagnostics: methods to assess Ensemble consistency and sensitivity, and methods to assess the reliability of Ensembles in probabilistic assimilation systems.
- Practical improvements to the design of Ensemble data assimilation schemes.
- Preparing future Ensemble DA for CMEMS.

The novelties for Ensemble generation of the proposed SP are:

- Performing Ensembles with the new available Copernicus Climate Change Service (C3S) Ensemble reanalysis dataset (ERA5; 10 members) and the European Center for Medium-Range Weather Forecasts Ensemble Prediction System (ECMWF-EPS; 50 members). The goal using multivariate atmospheric Ensembles is to obtain a large controlled (without DA) ocean Ensemble spread and focus on variables and time scales of particular interest to CMEMS.
- Complementing stochastic methods such as the Stochastic Parametrization of Perturbed Tendencies (SPPT; Brankart et al., 2015) and the Stochastic Perturbations of Unresolved Fluctuations (SPUF; Brankart, 2013). The SPUF method is investigated for the first time in the context of a high-resolution NEMO configuration (an identical subdomain of the IBI-MFC).

We will also perform complementary sensitivity studies:

- Inflation methods to produce less smooth Ensemble estimates (useful for the consistency analysis of Ensembles)
- Stochastic modelling of uncertainties in biogeochemical parameters (not only variables).
- Analysing the multivariate connections of physical-biogeochemical state variables, in support of Ensemble-based assimilation methods.

In addition, we have been in touch with Kristian Mogensen (ECMWF) about the possible use of prototype coupled ocean/atmosphere Ensemble members from ECMWF's forecast model, as an alternate reference for Ensemble consistency checking, and possibly as a way to enrich our base of uncertainties.

**3. Methodology and Experimental Design**

In this SP proposal, we will use a regional configuration of the NEMO community model (Nucleus for European Modelling of the Ocean; Madec, 2008; http://www.nemo-ocean.eu), BISCAY36, which has been set up in MyOcean, and has been used by the proponents (Quattrocchi et al., 2015; Vervatis et al., 2016; ECMWF-SP account id SPGRVERV). The model domain covers the Bay of Biscay and the western part of the English Channel, using a $1/36^{o}$ curvilinear Arakawa C-grid. For a complete description of the numerical set-up, identical to the IBI-MFC within CMEMS, and for details of validation, the reader is referred to Maraldi et al. (2013: also a LEGOS-MERCATOR Ocean collaboration).

The NEMO ocean engine OPA, in its latest version 3.6, is coupled with the passive tracer package TOP2 including the biogeochemical model PISCES-v2 (Pelagic Interactions Scheme for Carbon and Ecosystem Studies volume 2; Aumont et al., 2015). The meteorological fields will be provided by the C3S new ERA5 reanalysis dataset (10 members) and the operational ECMWF-EPS (50 members); the initial state and the components of the ocean-biogeochemical open boundary conditions will be inquired through the daily-weekly archives of the CMEMS infrastructure.

The experimental protocol with respect to the Ensemble forecast range (i.e. short/medium) is designed in a way that:

- The length of Ensemble time chunks will be adjusted between 2 weeks and a month, depending on early results
- Each time chunk contains in sequence, across the Ensemble:

- A short Ensemble spin-up period of a few days to a week, meant to bring the Ensemble up to empirical consistency with TAC data. The current SCRUM project shows that 50% of detectable array modes (above observational noise floor) characteristic of the Ensemble are consistent with OSTIA L4 data after about a week in the BISCAY36 configuration adopted.

- A Usable Period (UP) of "Ensemble forecasts", spanning the rest of the time chunk, meant to assess model uncertainties and probabilities over lead times within the length of the UP. Note that an analysis step will be needed during the UP for the Desroziers algorithm.

- The UPs will allow to study the dependence of the age of errors in the reliability diagnostics. We will also be able to properly compare different periods of the year since we can make the comparisons with identical ages of errors.

- An additional option will be to enrich the Ensembles by using samples from different periods with identical Lead Time (LT) and forecast range; in effect this method will multiply the size of Ensembles (the seasonal cycle and tidal signal are removed) in a similar way to phase errors in time-lagged Ensembles (Vervatis et al., 2016).

Those developments will be carried out in the same domain as in SCRUM (IBI-MFC subdomain), in two periods:

- **Period A** is identical to SCRUM for the years 2011-2012 in order to be able to compare the new developments with our previous results.
- **Period B** will be a period in the year 2017, aimed at checking the consistency of the new Ensembles compared to Sentinel-3A products, which are of particular interest to CMEMS.

We expect to generate Ensembles as in **Table 1**. Please note that we continue Ensemble numbering from the current SP numbering, which produces three Ensembles, Ens-1 to Ens-3. Those early SCRUM Ensembles will continue being used and validated in this project. In addition, the Ensembles of the ongoing and the new Special Project archived at ECMWF are considered as a community effort, to be shared with other interested groups within CMEMS SE, as well as in the broader community.

| Ensemble name | Contents |
|---|---|
| **Ens-4A** | SPPT-AR1, Period A |
| **Ens-4B** | SPPT-AR1, Period B |
| **Ens-5** | Several ensembles with ECMWF-EPS/ERA5 |
| **Ens-6** | Combined SPPT-AR1 + ECMWF-EPS |
| **S-** | Sensitivity tests (in the form of small Ensembles) |
| **C-** | Tests with small coupled ensembles |

**Table 1: Ensemble generation in this project**

## 4. Technical Requirements

The technical requirements for computational resources in this new Special Project follow closely those of the ongoing ECMWF SP with account id SPGRVERV, to end this year in 2018. The NEMO modelling configuration and the ensemble capabilities of the system is identical to the ongoing SP. In line with the "progress reports" and the "additional resources requests" published in https://www.ecmwf.int/en/research/special-projects/spgrverv-2016, we briefly summarize in the following paragraphs the resources and storage requirements for the ensemble experiments.

Sensitivity simulations have been carried out on CCA and CCB clusters, using the Intel Broadwell nodes with 36 physical cores and memory limit 120 MB. The optimal BISCAY36 domain decomposition is 96 CPU cores (excluding land points). The modelling system is compiled on Intel environment using the XIOS library to handle I/O processes. The experiments are designed for medium to seasonal range ensembles and between 20 to 50 members. An example of a batch job of 10 members on the Broadwell nodes consists of 96 * 10 = 960 CPU cores (hyperthreading = 1). The specific resources geometry fills 26 nodes (i.e. 26 * 36 = 936 CPU cores), plus 24 additional CPU cores on the 27th node. In order to fill properly the last node, we use 12 CPU cores for XIOS, dedicated only to I/O processes for the model's diagnostics (using the option to write in one netcdf file). Batch jobs are designed to be resubmitted every month of simulation, not exceeding a 18 hours walltime. For the latter example, the estimated resources and storage for 10 Ensemble members per month is about 200.000 SBUs and 300 GB (or 150 GB if we choose to store only the Chlorophyll classes from the ecosystem model for some of the Ensembles), respectively. For the first year of the project in 2018, we estimate (approximately after rounding numbers, taking also under account that this is a late request and simulations will start in the second quarter) at about 5 MSBUs resources and 7 TB disk storage. The allocated resources for the years 2019 and 2020 are estimated approximately at about 8 MSBUs and 2 MSBUs, respectively, reaching a total of 18 TB at the end of the project. Small deviations from the above numbers should be expected during the project, once we have all available information for the prototype coupled ensembles.

**References**


Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., Gehlen, M., 2015. PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies. Geosci. Model Dev. 8, 2465-2513.

Brankart, J.-M., Candille, G., Garnier, F., Calone, C., Melet, A., Bouttier, P.-A., Brasseur, P., Verron, J., 2015. A generic approach to explicit simulation of uncertainty in the NEMO ocean model. Geosci. Model Dev. 8, 1285-1297.

Brankart, J.-M., 2013: Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling. Ocean Modelling 66 (2013) 64–76, http://dx.doi.org/10.1016/j.ocemod.2013.02.004 .

De Mey, P., 2000: Scientific Feasibility of Data Assimilation in MERCATOR. MERCATOR report, reference MOO-ST-431-8-MER, Sept. 2000.

Huthnance, J.M., L.A. Mysak, and D.-P. Wang. 1986: Coastal trapped waves. In: "Baroclinic Processes on Continental Shelves", C.N.K. Mooers, ed., AGU, 1-18.

Kourafalou V.H., P. De Mey, M. Le Hénaff, G. Charria, C.A. Edwards, R. He, M. Herzfeld, A. Pasqual, E.V. Stanev, J. Tintoré, N. Usui, A.J. Van Der Westhuysen, J. Wilkin and X. Zhu, 2015. Coastal Ocean Forecasting: system integration and validation. Journal of Operational Oceanography, doi:10.1080/1755876X.2015. 1022336.

Kourafalou, V.H., P. De Mey, J. Staneva, N. Ayoub, A. Barth, Y. Chao, M. Cirano, J. Fiechter, M. Herzfeld, A. Kurapov, A.M. Moore, P. Oddo, J. Pullen, A.J. van der Westhuysen and R.H. Weisberg, 2015. Coastal Ocean Forecasting: science foundation and user benefits. Journal of Operational Oceanography, doi:10.1080/1755876X. 2015.1022348.

Madec, G., 2008. NEMO ocean engine. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27 ISSN No 1288-1619.

Maraldi, C., Chanut, J., Levier, B., Reffray, G., Ayoub, N., De Mey, P., Reffray, G., Lyard, F., Cailleau, S., Drevillon, M., Fanjul, E.A., Sotillo, M.G., Marsaleix, P., and the Mercator Team,

2013. NEMO on the shelf: assessment of the Iberia-Biscay-Ireland configuration. Ocean Sci. 9, 745-771.

Pinardi, N., L. Cavaleri, P. De Mey, C. Fratianni, I. Huthnance, P.F.J. Lermusiaux, A. Navarra, R. Preller, and S. Tibalidi, 2017 : From Weather to Ocean Predictions: an Historical Viewpoint. The Sea. Volume 17, The Science of Ocean Prediction, Part 1, Special Issue, J. Marine Res. 75(3). pp. 103-159. https://doi.org/10.1357/002224017821836789 .

Quattrocchi, G., De Mey, P., Ayoub, N., Vervatis, V.D., Testut C.E., Reffray, G., Chanut, J., Drillet, Y., 2014. Characterisation of errors of a regional model of the Bay of Biscay in response to wind uncertainties: a first step toward a data assimilation system suitable for coastal sea domains. J. Oper. Oceanogr. 7 (2), 25-34.

Vervatis, V., Testut, C.E., De Mey, P., Ayoub, N., Chanut, J., Quattrocchi, G., 2015. Data assimilative twin-experiment in a high-resolution Bay of Biscay configuration: 4D EnOI based on stochastic modelling of the wind forcing. Ocean Modell.,.

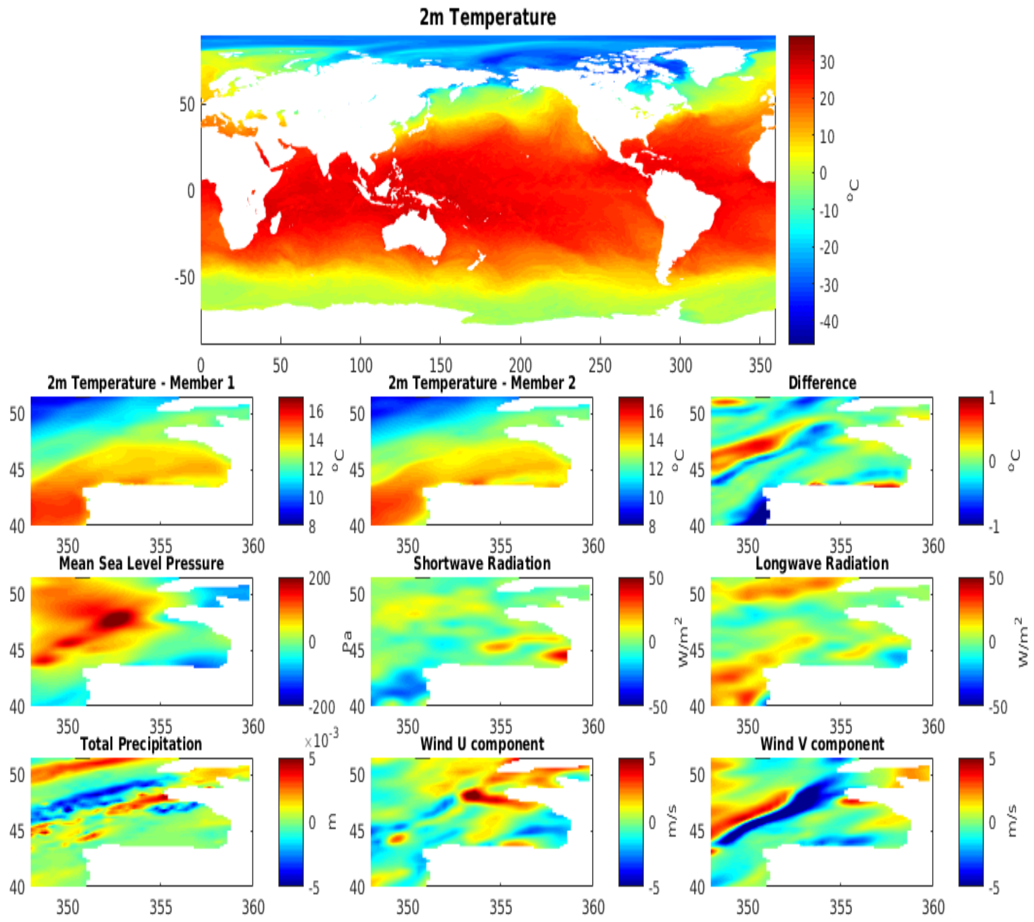


Fig. 1 ECMWF-EPS atmospheric Ensemble (date: 20120101); rows one-two: 2m air temperature in the global and Bay of Biscay domain for members 1 and 2 and their difference; rows three-four: difference of members 1 and 2 for other forcing variables.