SPECIAL PROJECT FINAL REPORT

Project Title:	Multi model monthly ensemble
Computer Project Account:	spitspia
Start Year - End Year :	2012 - 2014
Principal Investigator(s)	Valentina Pavan[1], Andrea Buzzi[2]
Affiliation/Address:	[1]Viale Silvani 6, 40122 Bologna, ITALY [2] Via Gobetti 101, 40129 Bologna, ITALY
Other Researchers (Name/Affiliation):	Daniele Mastrangelo, ISAC-CNR Piero Malguzzi, ISAC-CNR

The following should cover the entire project duration.

Summary of project objectives

The current project aims at producing calibrated multi-model ensemble monthly predictions over Italy and at comparing the multi-model system with single model system skill scores. The two systems used to produce the global ensemble monthly prediction are the GLOBO system by CNR, Bologna and the ECMWF IFS system.
Furthermore, within the framework of this project, it is done the work related to the productions of calibrated multi-model seasonal predictions over Italy obtained starting from the EUROSIP ensemble system and of their impacts on agriculture.

Summary of problems encountered

No problem was encountered

Experience with the Special Project framework

Our Special Project experience was positive.

Summary of results

See file attached.

List of publications/reports from the project with complete references

V. Pavan, F. J. Doblas-Reyes, 2013: Calibrated multi-model ensemble summer temperature predictions over Italy. Clim Dyn, 41, 2115-2132.

G. Villani, L. Botarelli, V. Marletto, A. Spisni, V. Pavan, W. Pratizzoli, F. Tomei, 2014: iCOLT – Seasonal forecasts of crop irrigation needs at ARPA-SIMC. ECMWF Newsletter, 138, 30-33.

Future plans

As described in the final report, the work done within this Special Project was crucial in order to obtain two major results: the participation of ISAC-CNR to the S2S WMO Project with the GLOBO system and the start of a new European Project focussed on probabilistic seasonal predictions of irrigation names MOSES(Managing Crop water Saving with Enterprise Services).

Final report of ECMWF Special Project SPIT-SPIA

D. Mastrangelo, P.Malguzzi, A. Buzzi (CNR-ISAC)
V. Pavan (ARPA-ER)

The SPIT-SPIA Special Project has been focused on producing multi-model ensemble operational predictions over Italy at seasonal to sub-seasonal time scale.
The work has been carried out on two separate issues: the implementation of an operational multi-model probabilistic monthly prediction system and the evaluation and production of multi-model calibrated probabilistic seasonal predictions of surface meteorological fields over Italy and of multi-model calibrated seasonal predictions of irrigation over Emilia-Romagna.
In the following, it is given a short overview of the results obtained and of the operational products built within the project, together with a description of their future expected applications.

GLOBO-IFS multi-model

Information about prediction skill is a crucial ingredient of a forecast product. This is especially true for extended range predictions, like probabilistic monthly forecasts, as they are affected by a greater rate of failure than their short to medium range counterparts. A method used to assess prediction skill at monthly time range is to run a set of reforecasts for several past dates using the same prediction system. Reforecasting is also an important step in long range forecasting because it is necessary to compute and remove the model systematic error.

The work done within the present Special Project was partly focused on building an example of multi-model application and model verification of probabilistic monthly predictions, using the CNR-ISAC prediction system, based on the atmospheric general circulation model GLOBO, and the ECMWF IFS monthly forecasting system.

In particular, the two systems have been used to evaluate the multi-model ability in predicting late-winter/early spring cold spells over Italy, focusing on the specific case of the cold spell which hit Emilia-Romagna in February 2012, associated with very heavy snowfall over the eastern part of the region (Grazzini, 2013).

During the course of the present project the GLOBO model developed at CNR-ISAC has been improved and exported onto the ECMWF computer facilities. GLOBO predictions are calibrated by means of a fixed set of reforecasts that covers the 30-year period from 1981 to 2010. This period constitutes the reference climate used to compute calibrated forecast anomalies in the CNR-ISAC operational forecasting system. The reforecast set is constructed by integrating the GLOBO model with initial dates differing by a fixed time interval of 5 days, starting from the 1st of January and ending the 27th of December of each year (with the only exception of leap years between the 25th of February and the 2nd of March). This gives a total of 73x30=2190 runs lasting 31 days each. Finally, for each new forecast date the GLOBO system is used to produce an ensemble forecast of 51 members initialized with analyses derived from the ECMWF-IFS ensemble forecast of that day.

The ECMWF-IFS monthly forecasting system includes a set of reforecasts performed "on the fly", on a weekly basis every Thursday, covering the past 20 years (Vitart, 2014). Besides the control run, the IFS reforecast set includes 4 ensemble members. Control and members are stored in the MARS ECMWF archive. Furthermore, the ECMWF-IFS monthly forecasts include 51 ensemble members for each new forecast date.

Despite the differences in the choices of initial dates between GLOBO and IFS, a quite large number of common dates can be found for the winter/early spring season, the period we focused on in this study. This season is defined here as the time interval that contains the reforecast runs with initial dates in the December-January-February (DJF) months. Over the 17 winters from 1994 to 2010 there exists an average of more than 14 cases per year (in 1992 and 1993 there are only few common dates) having coincident initial dates, giving a total of 249 runs per model that can be compared. This number is large enough to attempt a multi-model exercise.

For this purpose, the IFS control runs are downloaded from MARS for all common dates by selecting 500 hPa geopotential (Z500) and 850 hPa temperature (T850) as meteorological parameters. The fields are interpolated on a common verification grid of 1.0 degree and averaged in time over the first, second, third, and fourth week of forecast. A similar operation is performed for the GLOBO reforecast archive. For each grid point i, j, the multi-model prediction (MM) for week w =1,..,4 and initial date d=1,…,249 is defined as:

$$MM(i,j,w,d) = c_0(i,j,w) + c_1(i,j,w)M_1(i,j,w,d) + c_2(i,j,w)M_2(i,j,w,d) \quad (1)$$

where M1 and M2 denote the predictions of the GLOBO and IFS model, respectively. The weighting factors are determined by linear regression, i.e. by minimizing the cost-function

$$\sum_{d=1}^{249}(MM(i,j,w,d) - O(i,j,w,d))^2 \quad (2)$$

where O denotes the verifying observation, which in this study is taken from the ERA-Interim re-analyses. This is equivalent to the usual definition of multi-model (Krishnamurti et al., 2000, 2003) in terms of anomalies:

$$MMA = c_1(M_1 - \overline{M_1}) + c_2(M_2 - \overline{M_2}) \qquad (3)$$

with the overbar denoting the average over all initial dates, and with

$$c_0 = \bar{O} - c_1\overline{M_1} - c_2\overline{M_2} \qquad (4)$$

The weighting coefficients c_1 and c_2 are shown for Z500 (Fig. 1) and T850 (Fig. 2). Both figures show that the IFS model contributes more to the multi-model especially in the first and second week, whereas in the last two weeks the two contributions are similar, with areas with null or slightly negative values and maxima located sometimes in different zones. The different geographical distribution of the two models contributions is indicative of some degree of complementarity between them, a desirable property for models contributing to the same ensemble products.

Fig. 3 reports the sum of c_1 and c_2 for Z500, while Fig. 4 shows the same quantity but for T850. Although this quantity does not have a precise meaning, it gives relevant information concerning the geographical regions where predicted anomalies are affected by large errors. If the sum of the coefficients is close to zero then (see equation (4)) climate is the best prediction that can be made, at least in a statistical sense.

The mid-latitudes of Southern Hemisphere and Eastern Atlantic are characterized by little predictability in the third week and almost none in the fourth, especially for the temperature. Asia, Western Pacific, and Eastern America have some potential in the fourth week, as well as Antarctica and the Equatorial belt.

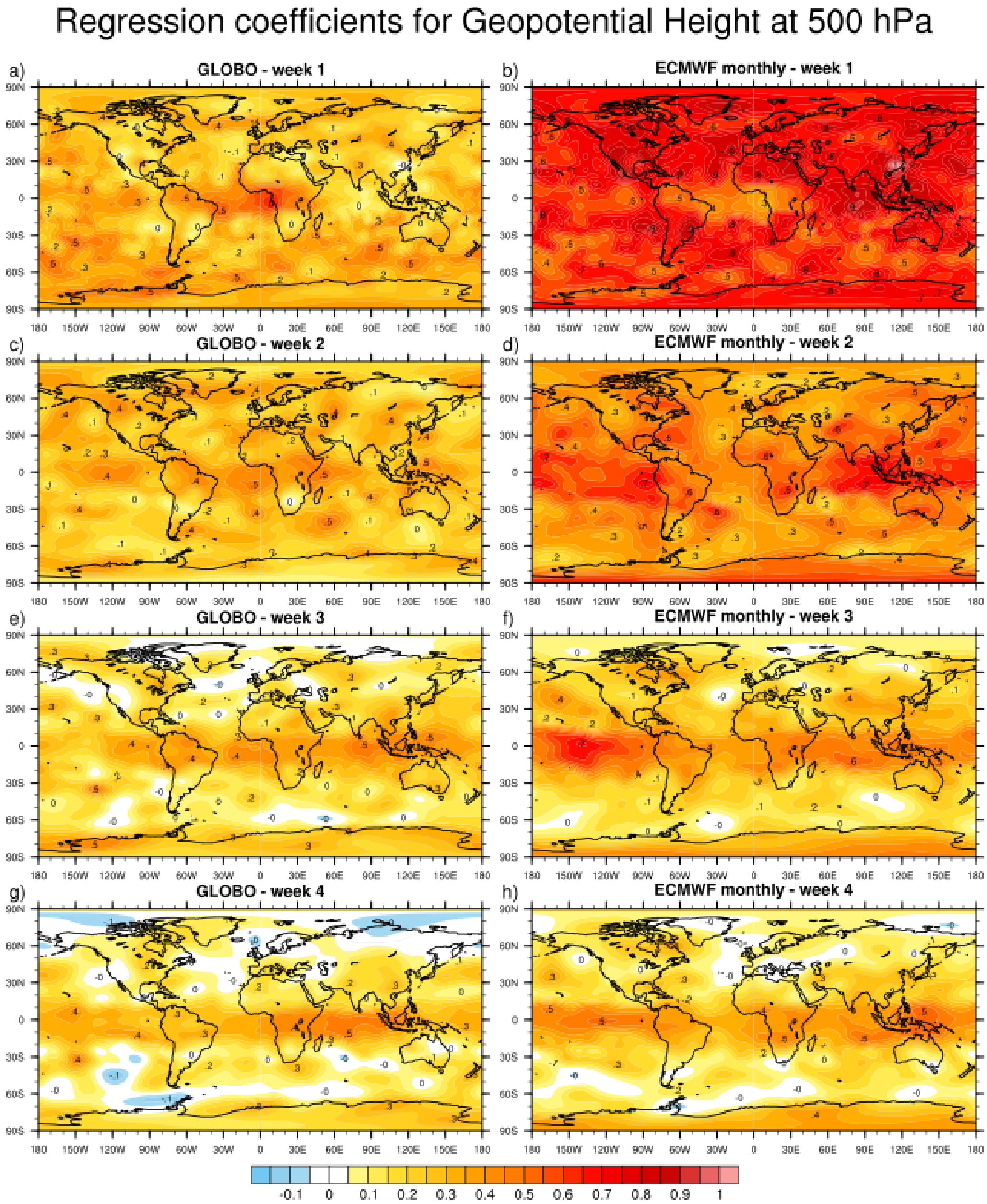

Fig 1: Regression coefficients for the GLOBO (left panels) and ECMWF monthly (right panels) models computed for the 500 hPa geopotential height.

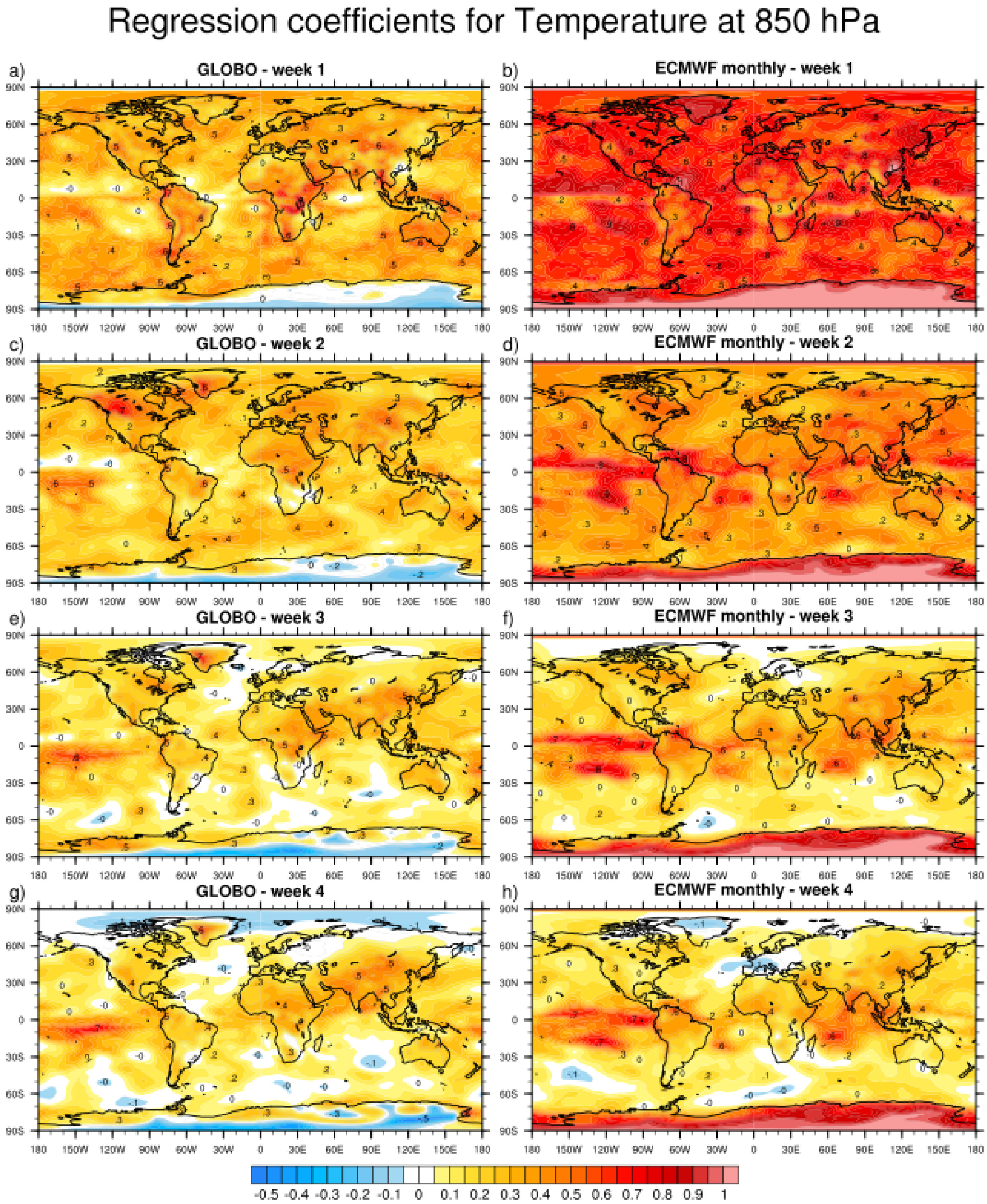

Fig 2: as in Fig 1, but for 850 hPa temperature

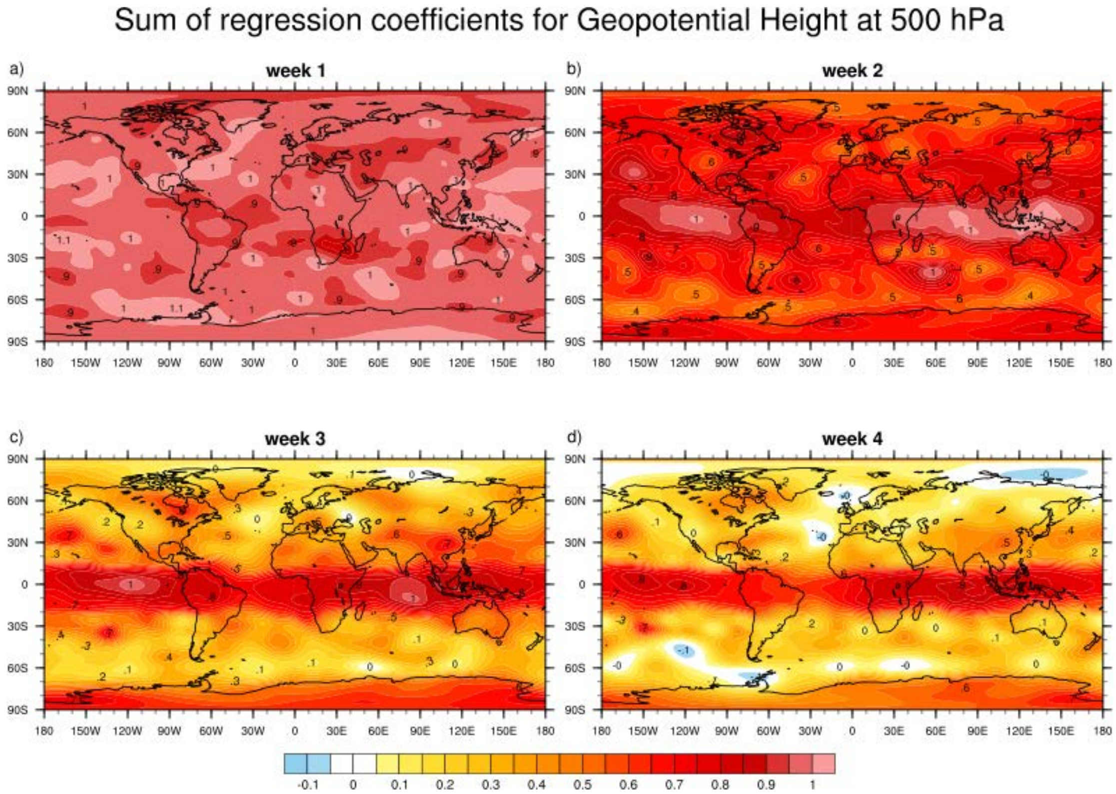

Fig. 3: Sum of the regression coefficients computed on a weekly basis for the 500 hPa geopotential height.

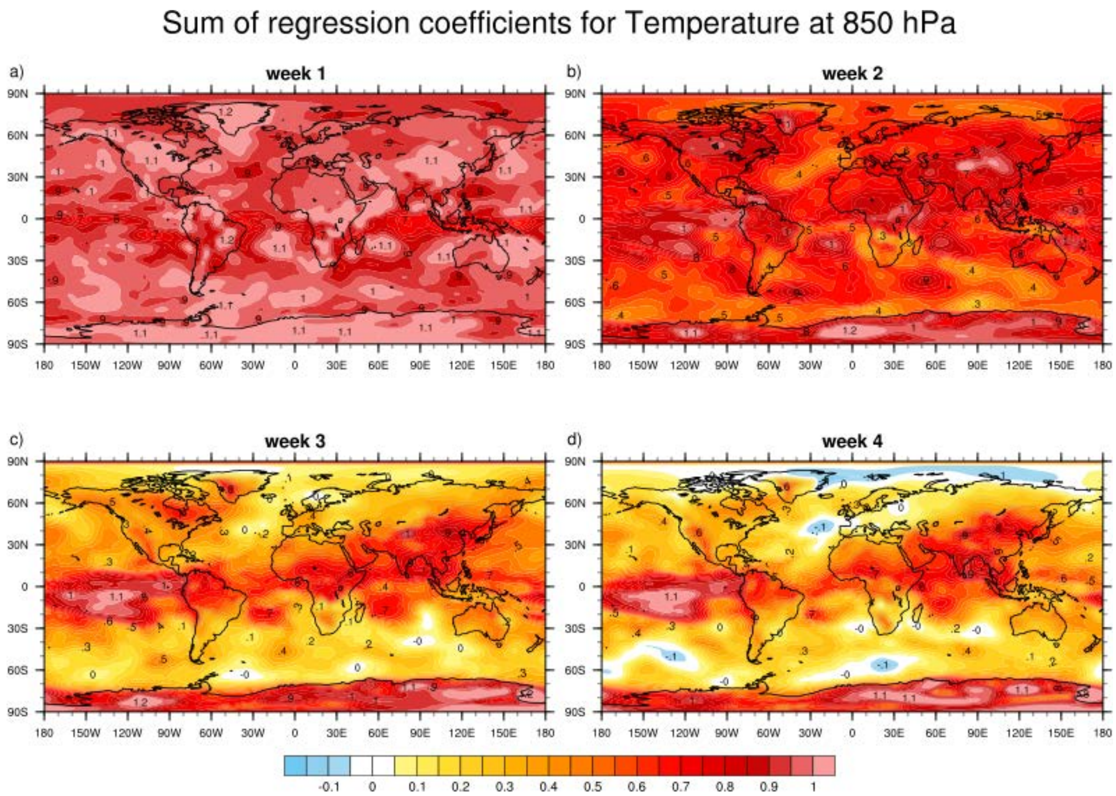

Fig. 4: As in Fig. 3, but for 850 hPa temperature.

Verification of multi-model performance

The set of reforecast runs with coincident initial dates can be also used for a statistical evaluation of the multi-model performance. To obtain an effective evaluation of the forecast skill, the entire set formed by 249 runs must be divided in a "training" and "verifying" subset. In this way, the multi-model prediction is verified on cases that do not affect the computation of the weighting factors (cross evaluation). To this end, the verifying subset is specified as a winter season, while the training set is made of all the remaining runs. The multi-model evaluation is then repeated for all available winters in a similar fashion. Since the first two winters do not contain enough cases, we omitted verification for those years (1992-1993), leaving a total of 17 winters over which average is performed.

Fig. 5 shows the results in terms of RMSE for various geographical areas and for Z500, while Fig. 6 reports the same results for T850. The anomaly RMSE of Fig. 5 indicates that the multi-model prediction performs significantly and systematically better than both models. Concerning the single-model RMSE, in general, the IFS shows slightly better results in the first and second week, while the GLOBO model outperforms IFS in the Southern Hemisphere in the third and fourth week. Same considerations can be obtained from Fig. 6. For the Equatorial belt, errors are typically smaller consistently with the smaller variance shown by the chosen atmospheric parameters in this region.

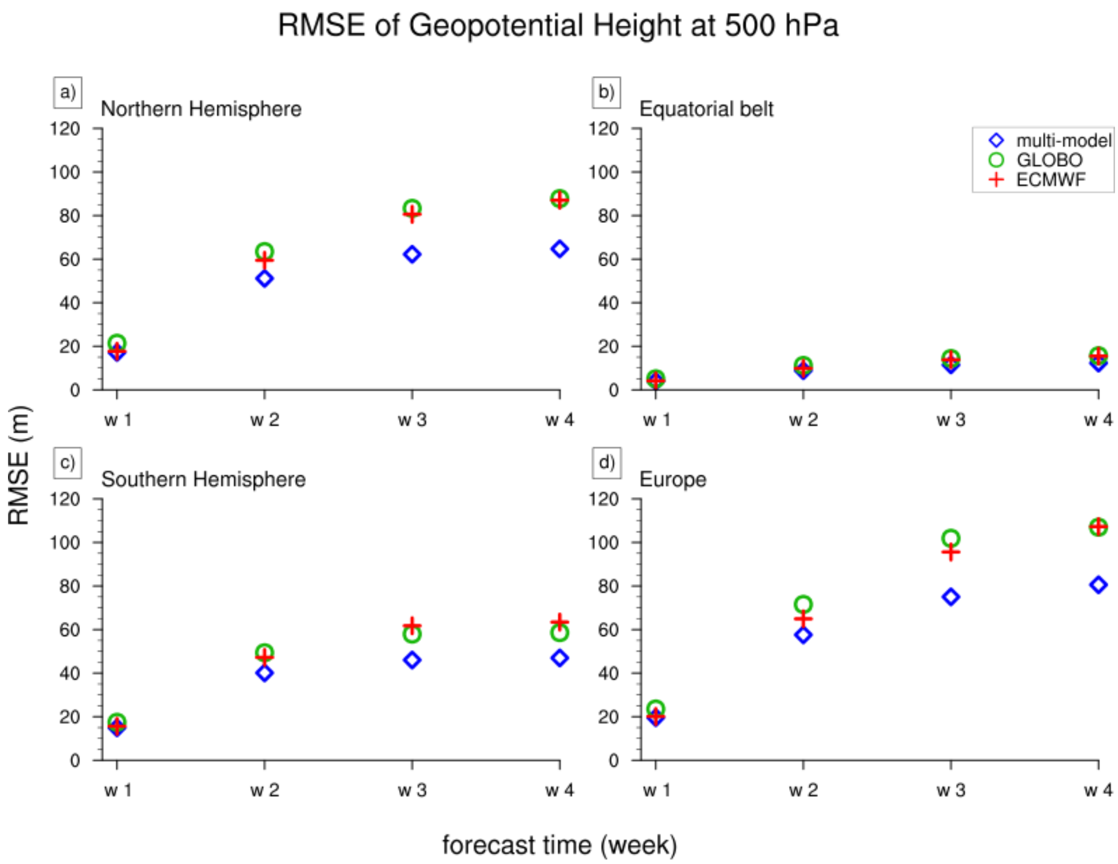

Fig 5: Root mean squared errors (RMSE) of the weekly averaged anomalies of 500-hPa geopotential height averaged over the selected 17 winters of GLOBO (red crosses) and ECMWF monthly model (green circles), and their multi-model combination (blue diamonds). The RMSE is computed over 4 different areas: a) the Northern extratropical hemisphere (20-90 N lat), b) the equatorial belt (-20 – 20 N lat), c) the Southern extratropical hemisphere (20-90 S lat), d) the European area (-20 – 60 E lon, 30 – 80 N lat).

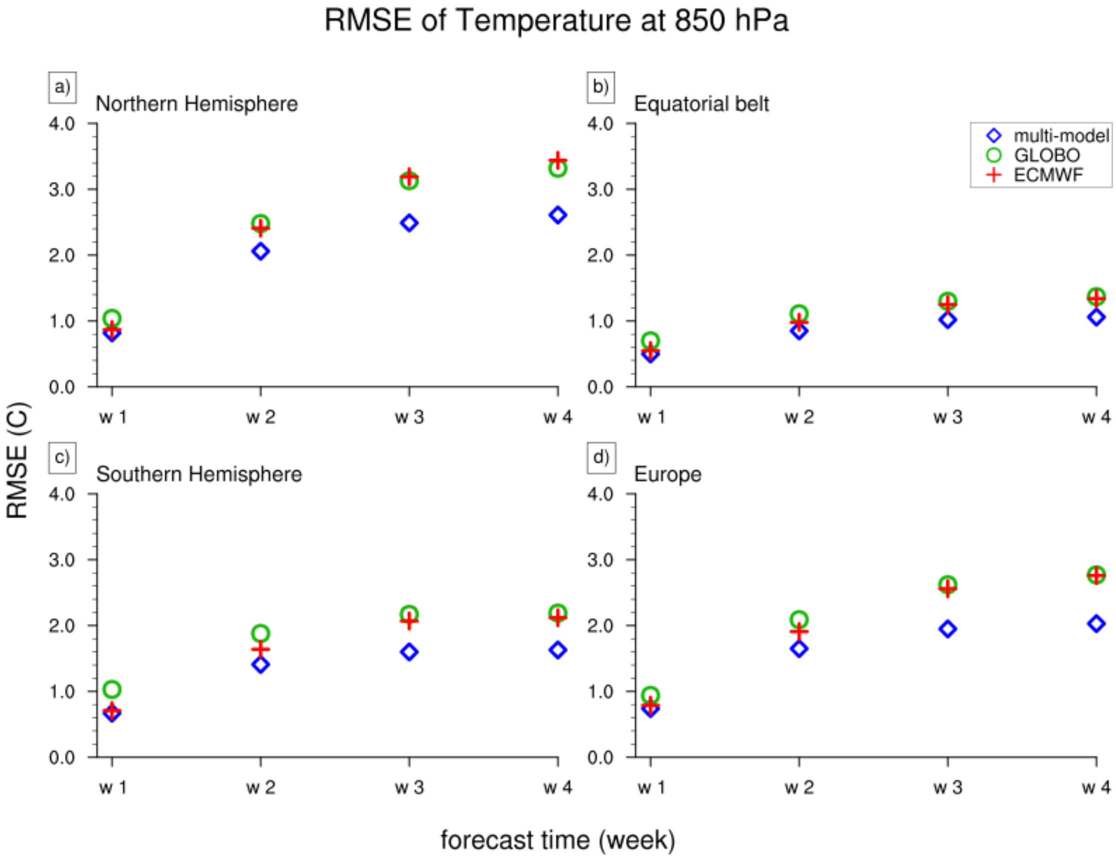

Fig 6: As in Fig 5 but for 850-hPa temperature.

Application to the case study of February 2012

The weighting coefficients have been tested on a case study, the February 2012 cold spell over Europe that produced extreme snowfall on Italy. The resulting multi-model forecast is compared to the two single-model forecasts: the GLOBO ensemble forecast and the ECMWF monthly ensemble forecast. Both forecasts are initialized on 00 UTC of 19 February 2012.

In general, the members that are at the basis of ensemble forecast differ essentially because small perturbations are added at the initial time. These perturbations grow in different direction of the phase space, so that member trajectories diverge typically beyond 2 weeks, consistently with chaos theory. This is also true for control runs, since there is no substantial difference between control and members of the same model. The weighting factors that have been computed with control runs of the IFS and GLOBO models must henceforth be valid also for each pair of members of the same model, and consequently for averages of members, due to the linearity of the mean operator. Therefore, the weights that minimize the cost function for control runs can be also used for ensemble averages.

Results from the three forecasts (GLOBO, ECMWF and multi-model ensembles) are shown for the 4th week in Fig. 7 and 8 for Z500 and T850, respectively.

Panels c of both figures show that the multi-model anomalies are weaker than their single-model counterparts (panels a and b), as a consequence of averaging. The negative Z500 anomalies forecasted by the two models in different areas of the Northern Pacific contribute to produce the strongest Z500 negative anomaly of the multi-model forecast. Indeed, the latter is collocated with a local maximum of the sum of the coefficients shown in Fig. 3d. The same happens for the positive Z500 anomaly over eastern Canada too. On the other hand, the negative anomaly centered over the British Islands in the two single-model forecasts, which has no observational counterpart, is nearly filtered out in the multi-model forecast.

As for T850, the multi-model ensemble mean features a negative anomaly over north-western Pacific and north-eastern Asia, and a positive anomaly over North America. Over Europe the signal is weak and the ensemble mean predicted anomaly is mostly on opposite phase with respect to observations.

The resulting weak anomalies forecasted for the 4th week are evaluated in terms of the deterministic scores shown in Tab1, for Z500, and Tab. 2, for T850. For both the fields, the multi-model prediction represents an improvement with respect to single-model predictions, both in terms of RMSE and ACC, averaged on the whole extratropical Northern Hemisphere.

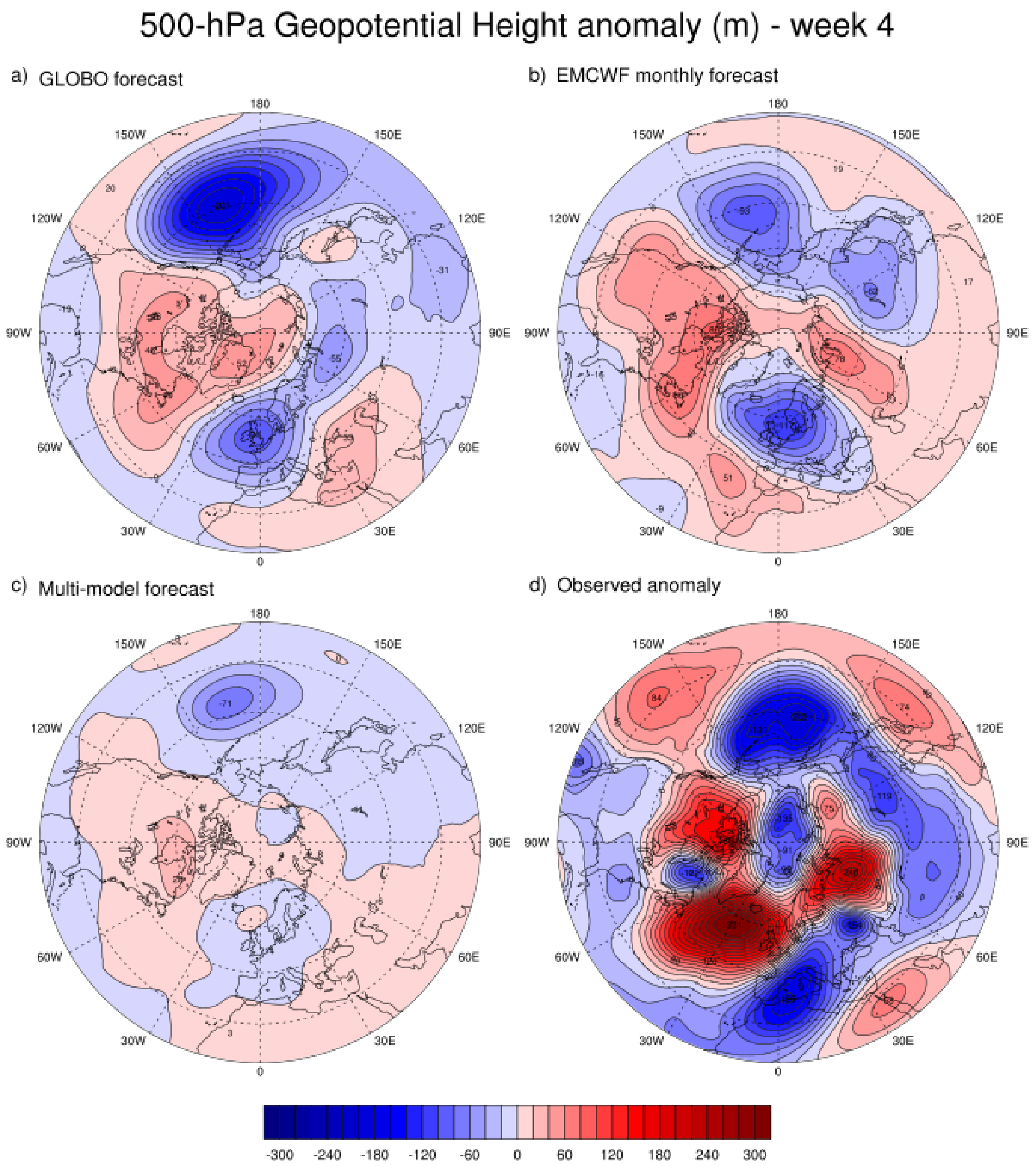

Fig. 7: 500 hPa geopotential height anomalies (m) averaged on the fourth week of the forecast initialized on 23 January 2012 with the GLOBO model (a), the ECMWF monthly forecasting system (b), the multi-model ensemble (c). The verifying anomaly (d) is computed from ERA-Interim re-analyses.

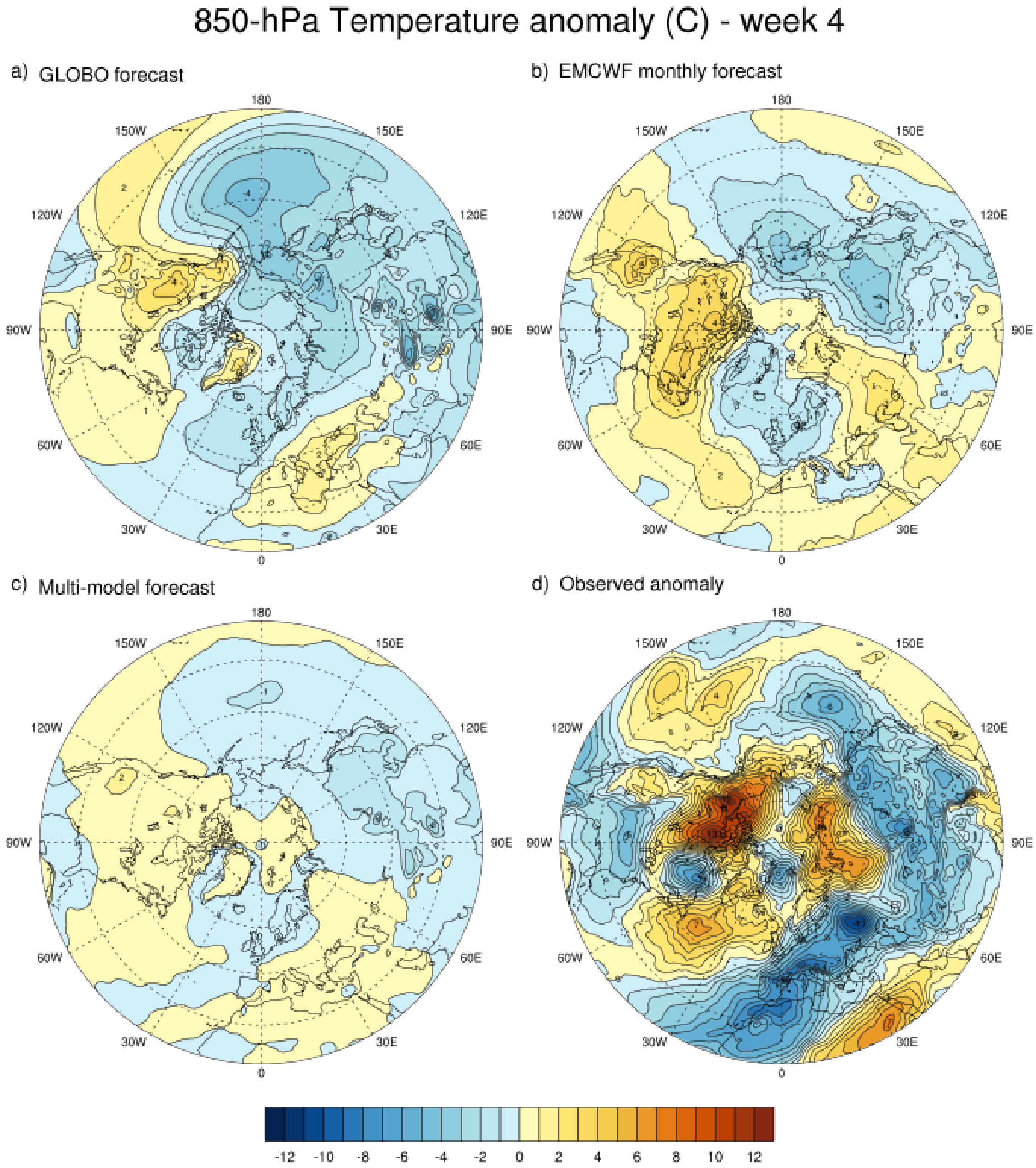

Fig. 8: As in Fig. 7 but for 850 hPa temperature.

Week 4	RMSE-NH (m)	ACC NH
Multi-model	86.1	0.30
GLOBO	93.1	0.16
ECMWF Monthly	87.0	0.26

Tab. 1: Root Mean Squared Error (RMSE) and Anomaly Correlation Coefficient (ACC) of 500 hPa averaged on the 4th week of the multi-model, GLOBO, and ECMWF monthly forecasts.

Week 4	RMSE-NH (C)	ACC NH
Multi-model	3.61	0.27
GLOBO	3.99	0.01
ECMWF Monthly	3.64	0.25

Tab. 2: As in Tab. 1 but for 850 hPa temperature.

Multi-model ensemble monthly prediction discussion and Conclusions

This multi-model application on a single case study clearly illustrates how the multi-model works: the filtering implicit in this technique dumps those features that are not statistically correlated with the anomalies observed in the training period, retaining and slightly improving the skill (RMSE) that was present in the best model.
The results shown here in terms of ensemble mean and deterministic scores do not catch, at the end of the forecast, the cold spell whose intensity was the basic motivation for the choice of this case study. This result is not unexpected since, as already pointed out by Thompson in 1977, the prediction of extremes cannot be tackled with this deterministic-like application of the multi-model technique.
The use of probabilistic ensemble reforecasts, and its evaluation on many more test cases, should be adopted to produce a multi-model probabilistic forecast that could improve the extended-range forecast of categorical events as, for example, the probability of the lowest temperature terciles. Nonetheless, to this end, the computation of the multi-model regression coefficients is still a preliminary step that can lead to more reliable categorical probability estimates, for instance through logistic regression (Whitaker et al., 2006).
The CNR-ISAC operational system used within this Special Project is currently employed to contribute to the Expert Group for Monthly to Seasonal Predictions over Italy, coordinated by the National Civil Protection Agency and aimed at mitigating the impacts of intense climate anomalies over the Italian territory. The GLOBO forecasting system has also been included in the Sub-seasonal to Seasonal Prediction Project (S2S) recently promoted by WMO within the WCRP and WWRP Programmes (Robertson et al., 2015). Multi-model ensemble prediction at time ranges from monthly to seasonal is one of the main objectives of S2S. Therefore, the multi-model exercise developed in this Special Project gives a contribution to the development of these ongoing activities.

Seasonal prediction products

During the present Special Project it was possible to finalize the multi-model calibrated operational products built at ARPA-SIMC and obtained by applying a MOS empirical calibration and downscaling scheme based on Multi Linear Regression to EUROSIP multi-model operational seasonal predictions. In this way, high resolution calibrated statistical seasonal predictions were operationally produced for several climate indices over Italy, based on daily precipitation and daily minimum and maximum temperature. These predictions represented the main contribution of ARPA-SIMC to the Expert Group for Monthly to Seasonal Predictions over Italy coordinated by the National Civil Protection Agency.
The skill of this system has been assessed over the summer and for maximum temperature over Italy by Pavan and Doblas-Reyes (2013) by applying the scheme to the ENSEMBLES multi-model seasonal predictions DMO. The results show the superiority of the multi-model approach with respect to a single-model one and highlight the ability of the MOS scheme to reduce substantially the impacts of systematic model errors on the final forecast over Italy.
During the course of the Project ARPA-SIMC also developed a system able to produce probabilistic multi-model seasonal predictions of irrigation water need over Emilia-Romagna. The system exploits the aforementioned calibrated probabilistic multi-model summer predictions as input of a system consisting of a weather generator, a model of the surface water table level, and a water balance model (CRITERIA). The system produces, as output, water irrigation needs for the next summer season. Descriptions of the operational system and of some preliminary results obtained during the first three years of operational activity have been presented in Villani et al. (2014). These results indicate a greater skill of the irrigation water need seasonal prediction system with respect to the original DMO or calibrated seasonal products. It is thought that the better performance of the final irrigation product is partly linked to the relevance of a good initialization of the water table depth and partly linked to a good evaluation of the geographical distribution within the region of crop classes, characterized by different water needs.
Finally, the irrigation water need seasonal prediction system has been included as a core product in the new European Project MOSES (Managing Crop water Saving with Enterprise Services) started in July 2015. MOSES aims at integrating this prediction system with other agronomic monitoring systems based on satellite products and exploiting extended–to–monthly range probabilistic predictions. All of them will be applied within the MOSES project to several pilot areas within Europe, creating a starting point for a new European climate service targeted for agronomical applications.

References

F. Grazzini, 2013: Cold spell prediction beyond a week: extreme snowfall events in February 2012 in Italy. ECMWF Newsletters No. 136, 31-35.

T. N. Krishnamurti, C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel Ensemble Forecasts for Weather and Seasonal Climate. J. Climate, 13, 4196–4216.
doi: http://dx.doi.org/10.1175/1520-0442(2000)013<4196:MEFFWA>2.0.CO;2

T. N. Krishnamurti, K. Rajendran, T. S. V. Vijaya Kumar, S. Lord, Z. Toth, X. Zou, S. Cocke, J. E. Ahlquist, and I. Michael Navon, 2003: Improved Skill for the Anomaly Correlation of Geopotential Heights at 500 hPa. Mon. Wea. Rev., 131, 1082–1102.
doi: http://dx.doi.org/10.1175/1520-0493(2003)131<1082:ISFTAC>2.0.CO;2

V. Pavan, F. J. Doblas-Reyes, 2013: Calibrated multi-model ensemble summer temperature predictions over Italy. Clim Dyn, 41, 2115-2132.

P. D. Thompson, 1977: How to Improve Accuracy by Combining Independent Forecasts. Mon. Wea. Rev., 105, 228–229.
doi: http://dx.doi.org/10.1175/1520-0493(1977)105<0228:HTIABC>2.0.CO;2

A. W. Robertson, A. Kumar, M.s Peña, and F. Vitart, 2015: Improving and Promoting Subseasonal to Seasonal Prediction. Bull. Amer. Meteor. Soc., 96, ES49–ES53. .doi: http://dx.doi.org/10.1175/BAMS-D-14-00139.1

G. Villani, L. Botarelli, V. Marletto, A. Spisni, V. Pavan, W. Pratizzoli, F. Tomei, 2014: iCOLT – Seasonal forecasts of crop irrigation needs at ARPA-SIMC. ECMWF Newsletter, 138, 30-33.

F. Vitart (2014), Evolution of ECMWF sub-seasonal forecast skill scores. Q.J.R. Meteorol. Soc., 140: 1889–1899. doi: 10.1002/qj.2256

J. S. Whitaker, X. Wei, and F. Vitart, 2006: Improving Week-2 Forecasts with Multimodel Reforecast Ensembles. Mon. Wea. Rev., 134, 2279–2284.
doi: http://dx.doi.org/10.1175/MWR3175.1