

R_MEFH_MEDISEH routine for estimating nursery and spawning grounds of demersal species (version 2.0)

[Angela Martiradonna](#), Walter Zupa, Isabella Bitetto, Maria Teresa Facchini,
Maria Teresa Spedicato, Giuseppe Lembo

COISPA Tecnologia & Ricerca, via dei Trulli 18-20, Bari (Italy)
February 2013

The R_MEFH_MEDISH routine (Modelling Essential Fish Habitat, version 2.0) is written in R language and developed with R 2.15.2. It allows to estimate annual density maps along with annual and persistent hot spots for fish populations, starting from the abundance indices datasets extracted from the JRC data base using the R_RSI_MEDISEH R routine (Facchini *et al.*, 2012).

The annual densities of recruits and spawners are modeled as a function of depth with a ZIGAM approach (Liu and Chan, 2009) and, in order to capture the spatial correlation structure, ordinary kriging is performed on ZIGAM residuals. Assuming a rather stable spatial correlation structure throughout years, a pooled variogram (Morfin *et al.*, 2012) is estimated and the cross validation index VSE (Variance of Standard Error – Bleines *et al.*, 2000) is computed. Annual abundance maps are produced by predicting the ZIGAM and kriging models on a grid with the resolution of 1 km² generated inside a specific polygon, accounting for the known species distribution.

The Getis and Ord statistic (G statistic; Getis and Ord, 1992) is computed on the predicted values and hot spots could be identified with two approaches:

- Getis and Ord test with level of significance 0.95 (Getis and Ord, 2001): the hot spot is composed by all points of the grid whose standardize G statistic value is greater than the quantile of order $1 - \frac{0.95}{2N}$ of the standard normal distribution, where N is the amount of predicted values on the grid. The script `localG.r` (Mastrantonio, 2012) was used for this issue.

- Fisher homogeneous classification: the values of the statistic are classified in three classes such that the variance within groups is minimized and the highest class is selected as a nursery or spawning ground. For this purpose the Fisher algorithm (Fisher, 1958) is used.

Annual abundance estimated values are mapped and hot spots are highlighted with a red contour.

The annual abundance maps are reclassified attributing the value 1 to the points of the grid belonging to an estimated hot spot and 0 to the others. The reclassified binomial maps are overlapped and the persistence indices are computed by arithmetical mean of 0-1 occurrence. The results, partitioned in five classes (0.05-0.20, 0.21-0.40, 0.41-0.60, 0.61-0.80 and 0.81-1), are plotted in the persistence map.

1. Before running R_MEFH_MEDISH

Before running R_MEFH_MEDISH the first time it is necessary to install R 2.15.2 (available at <http://cran.r-project.org/>) and all the libraries stored in the package directory.

Input data

For each GSA a depth grid is required as a .csv file of the type

lon	lat	lon_utm	lat_utm	depth
18.31395	42.40328	772.728	4699.875	178
18.32558	42.40328	773.6851	4699.913	166
18.33721	42.40328	774.6421	4699.95	159
18.34884	42.40328	775.5991	4699.988	151
...
...

where:

- **lon** and **lat** are the coordinates measured in decimal degrees;
- **lon_utm** and **lat_utm** are the UTM projected coordinates (km).

That grid should be stored in the `grids` directory.

For each case study (GSA/species/life stage) the abundance indices dataset provided by `R_RSI_MEDISEH` is required and it should be stored in the `datasets` directory.

2. Run `R_MEFH_MEDISH`

The routine is structured in 7 main steps.

STEP 1 - Settings

The user should set:

- the GSA number;
- the UTM zone for coordinates projection;
- the species MEDITS code in capital letters (e.g. “PAPE LON”);
- the life stage, choosing between “*recruits*” or “*spawners*”;
- the name of the dataset provided by `R_RSI_MEDISEH` and stored in the `datasets` directory
(e.g. “PAPE LON10_1994-2010_metaDB recruits abundance indices.csv”);
- the name of the gsa depth grid stored in the `grids` directory
(e.g. “1km grid 0-800 GSA10.csv”).

STEP 2 – Initialization

Libraries are loaded, data are read and coordinates are projected in UTM system. GSA extreme coordinates are read from the *gsa coordinates.csv* file stored in the `grids` directory.

An output folder, named as `Results - GSA10 PAPE LON recruits`, where all results are saved, is generated.

Bubble plots of abundance indices are generated and saved in the `bubble plot` directory.

After looking bubble plots, the user should select the years in which do not perform the analysis by indicating, for example:

- `yrs_no_analysis <- c(1994,1995)` to exclude years 1994 and 1995;

- `yrs_no_analysis <- c()` to perform the analysis in all year in the time series.

Subsequently a histogram of depth distribution of positive hauls is generated which could be useful to locate the depth distribution of the population. The user should then indicate the minimum and the maximum value of depth in order to select the bathymetric polygon in which analysis will be carried out. Finally the bathymetric grid is loaded.

STEP 3 - Pooled variogram estimation

The user should first indicate the years to eventually exclude and the maximum distance (km) for pooled variogram estimation.

Running the line

source("scripts\\variogram.r"),

a window will appear with the pooled variogram graph and three variogram models (exponential, spherical and gaussian) fitted on it. The user could vary the number of lag-classes until he is satisfied of the fitting. He could:

- print “+” for increase the number of lag-classes;
- print “-“ for reduce the number of lag-classes;
- indicate a specific number of lag classes;
- print “ok” for accept the result;
- print “no” for exit.

After printing “ok”, the user should indicate the variogram model, choosing among “exp”, “sph” and “gaus”.

A summary table of variogram estimation is saved and the cross validation index (SVE) is computed.

STEP 4 - ZIGAM analysis

For each year a ZIGAM model with unique covariate *Depth* is estimated. The user should indicate the years in which do not perform ZIGAM analysis and the initial degrees of

freedom (K) for the smoothing function. Model graphs and summaries are stored in the `model` directory.

STEP 5 - Interpolation on residuals

If pooled variogram estimation (STEP 3) was successful carried out, a kriging interpolation on residual could be performed. The alternative solution is Inverse Distance Weighted (IDW) interpolation. The user should therefore select the interpolation type between “*kriging*” and “*idw*”. The default power coefficient for IDW interpolation is 5.

STEP 6 - Hot spots identification and mapping

The user should indicate the radius for G statistic computation and choose the hot spot identification criterion between “*fisher*” (for Fisher homogeneous intervals method) and “*getis_ord*” (for Getis and Ord test). The default number of classes for Fisher algorithm is 3 and the level of confidence for Getis and Ord test is 0.95.

The neighborhood net for G statistic computation is generated only the first time for each GSA and bathymetric extension and saved in the `grids` directory as

*neigh 10-800 5km gsa18.Rdata*¹.

Annual abundance maps with hot spots contours are generated and saved in the `maps` folder.

STEP 7 - Persistence analysis

Persistence indices are computed and classified in 5 classes:

CLASS 1: from 0.81 to 1

CLASS 2: from 0.61 to 0.80

CLASS 3: from 0.41 to 0.60

CLASS 4: from 0.21 to 0.40

CLASS 5: from 0.05 to 0.20

¹ If you change depth grid remember to delete the correspondent neigh net!

Polygon shape file are generated by mean the script `XYpoints2polygShapes.r` by M. Pilar Tugores (2012).

3. Other outputs

For each year a .csv file of the type

GSA	Species	Lear	Life stage	Prevision (num/km2)	X	Y	hot spot
18	PAPE LON	2010	R	65.126	18.53488	42.43934	0
18	PAPE LON	2010	R	64.3779	18.54651	42.43934	0
18	PAPE LON	2010	R	94.33514	18.4186	42.42131	0
18	PAPE LON	2010	R	89.43503	18.43023	42.42131	0
18	PAPE LON	2010	R	89.15799	18.44186	42.42131	1
18	PAPE LON	2010	R	84.583	18.45349	42.42131	1
...

is generated and saved in the specific folder with the name of the type
GSA18 PAPE LON recruits.csv.

Moreover a modeling table of the type

Species	Life Stage	Gsa	Depth Range	Period	Model Type	Covariates	Explained Variance (%)	Goodness of fit (R2)
PAPE LON	recruits	18	10-800	1996	Gaussian GAM	Depth	32.5	0.27
					Binomial GAM	Depth	28	0.24
					Ordinary Kriging (*)			
PAPE LON	recruits	18	10-800	1997	Gaussian GAM	Depth	7.52	0.02
					Binomial GAM	Depth	31.5	0.24
					Ordinary Kriging (*)			
...

is generated and saved in the main results folder.

In the end a summary table with the estimated parameters of each case study in the same GSA is produced and saved in the root folder as “*SUMMARY TABLE gsa18.csv*” (Table 1).

Table 1 - SUMMARY TABLE gsa18

GSA	Species	Life stage	Years	Zigam	Pooled Variogram	Kriging	Model	Range	Sill	Nugget	Cross valid VSE	IDW	Power coef	Hotspots	Neigh radius	Num fisher classes	Conf level
18	MERL MER	recruits	1996	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	1997	no	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	1998	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	1999	no	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2000	no	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2001	no	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2002	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2003	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2004	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2005	no	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2006	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2007	no	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2008	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2009	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	MERL MER	recruits	2010	yes	yes	yes	gaus	12.25	1291086	275146.8	0.69	no	-	fisher	5	3	-
18	PAPE LON	recruits	1996	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	1997	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	1998	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	1999	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2000	no	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2001	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2002	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2003	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2004	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2005	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2006	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2007	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2008	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2009	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-
18	PAPE LON	recruits	2010	yes	yes	yes	sph	17.85	55992.32	877.22	1.81	no	-	fisher	5	3	-

References

- Bleines C., S. Perseval, F.Rambert, D. Renard, Y. Touffait (2000) – ISATIS. Isatis Software Manual. Geovariance & Ecole des Mines de Paris, pp. 585.
- Facchini M.T., I. Bitetto, M.T. Spedicato, G. Lembo (2012) – MEDISEH (MEDIterranean Sensitive Habitats) – R routine for the estimation of mean length of Recruits and Spawners and for the *calculation of Indices from scientific survey data*. Specific project Mediseh, Marea Framework, mareaproject.net.
- Fisher W.D. (1958) – *On grouping for maximum homogeneity*. American Statistical Association Journal, December 1958: 789-798.
- Getis A., J.K. Ord (1992) - *The Analysis of Spatial Association by Use of Distance Statistics*. Geographical Analysis, Vol. 24, No. 3: 189-206.
- Getis A., J.K. Ord (2001) – *Testing for local spatial autocorrelation in the presence of global autocorrelation*. Journal of Regional Science, Vol. 41, No. 3, 2001, pp. 411-432.
- Liu H., K.S. Chan (2009). *Constrained Generalized Additive Models for Zero-Inflated Data*. Technical report #388 of Department of Statistics & Actuarial Sciences, University of Iowa.
- Martiradonna A., W. Zupa, I. Bitetto, M.T. Facchini, M.T. Spedicato, G. Lembo (2012) – MEDISEH (MEDIterranean Sensitive Habitats) – *R_MEFH_MEDISEH routine for estimating nursery and spawning grounds of demersal species*. Specific project Mediseh, Marea Framework, mareaproject.net.
- Mastrantonio G. (2012) - MEDISEH (MEDIterranean Sensitive Habitats) - mareaproject.net.
- Morfin M, Fromentin J-M, Jadaud A, Bez N (2012) - *Spatio-Temporal Patterns of Key Exploited Marine Species in the Northwestern Mediterranean Sea*. PLoS ONE 7(5): e37907. doi:10.1371/journal.pone.0037907.