Appendix III

Scale-dependent spatial modelling of the distribution of a marine predator: fin whale distribution in the Bay of Biscay

G. Certain, V. Ridoux & O. Van Canneyt

University of La Rochelle, Centre de Recherche sur les Mammiferes Marins, Avenue du Lazaret, 17000 La Rochelle, France

INTRODUCTION :

Recently, there has been a huge amount of scientific literature focusing on marine predator spatial distribution, with particular attention paid on the spatial scale(s) at which it was structured (Schneider & Piatt 1986 ; Russel *et al.* 1992 ; Fauchald *et al.* 2000). Spatial structure of animal population is one key parameter to understand how animals distribute in space and how they interact with the various components of their surrounding environment. The theory of hierarchical patch dynamics (Kotliar & Wiens 1990) has recently proposed an ecological framework to study the spatial distribution of animals. It states that the spatial distribution of animal populations can be viewed as a succession of nested patches, with small patches of high density located into larger patches of smaller density. Each patch is issued from the interaction between animals and environmental parameters structured at the same spatial scale than the patch.

In this study, our aim is to use CODA data in order to investigate the spatial structure of the distribution of Fin Whales (FW) *Balaenoptera physalus*, a species commonly encountered in the pelagic area of the Bay of Biscay which is also of strong conservation interest. There is an important benefit to identify the oceanographic factors determining the spatial structure of FW populations, to better understand the ecological processes affecting their spatial distribution in order to *in fine* achieve accurate spatial prediction of FW distribution at various dates and at relevant scale(s) for scientists and managers. To this end, we used a set of oceanographic covariates, each being spatially structured at different scales.

First, the spatial scale at which the population of FW was structured in the Bay of Biscay was investigated using spatial correlograms (Bjornstadt & Falck 2001). Second, the spatial distribution of FW was filtered at each scale with filtering kriging (Wackernagel *et al.* 2003). Filtering kriging consist in a spatial decomposition of the data that provides for a given dataset a number of variables (or filters) that describes the spatial distribution of the species at a given scale. In a third step, various covariates were extracted from satellite-derived products, and their associated gradients were calculated. Again, using spatial correlograms, the scale at which each environmental variable was structured has been calculated. Endly, each spatial filter was modelled using Generalised Additive Models (Wood & Augustin 2001) and environmental covariates spatially structured at an equivalent scale, in order to obtain one spatial predictive model relevant from each spatial scale at which FW populations were structured. These models were used to identify which environmental parameters influenced FW spatial distribution at which scales, and outputs of these models shows the areas where environmental parameters are the most suitable for FW.

DATA ANALYSIS:

Pre-processing the FW dataset:

The dataset we used was constituted of Fin Whale observations carried out during CODA cruises by primary observers only. In a first time, transects were sliced in a succession of small sized segments (0.1 decimal degree long, ~5.5km), each containing a number of observed FW individuals. Then a correction of segmented data has been carried out. This correction has been achieved using distance sampling methods, with a hazard rate model including factor sightability and observation platform height as covariates.

Corrected data were carefully examined in order to determine whether or not a segmented, corrected dataset was suitable for inclusion in a spatial model of species relative density. The main subject of our study is to understand which environmental factors influence FW distribution, in order to identify the areas where these factors are the most suitable for FW concentration. To do so, any strong distortion of the original dataset should be avoided, so that the modelled species–environment relationships truly reflect ecological mechanisms, as far as possible. Hence, the use of corrected segmented data is possible if such data are biologically meaningful. Therefore, corrected data were evaluated according to two parameters: the maximum density of FW per segments and the spatial structure existing in FW corrected data. We assume that a "biologically meaningful" correction should not affect strongly the spatial structure of the segmented data, and that maximum FW densities should remains relatively low, as FW are not known to form huge aggregations of individuals at the same location at sea.

Analysing the spatial structure of FW populations:

We used spatial correlograms to test for spatial autocorrelation in the FW dataset. Correlograms were computed with the NCF package (Bjornstadt 2006), available at http://asi23.ent.psu.edu/onb1/, and using R software (R development core team 2006). In addition, experimental spatial variograms were also computed with the gstat package (Pebesma & Wesseling 2003) to get additional information on the scales at which the FW dataset was spatially structured, and a variogram model was fitted to this experimental variogram. The adjustment criterion of the variogram model was weighted least squares, where the weights are proportional to the number of point pair for each distance lag. Once relevant spatial scales were identified, spatial filters were extracted with filtering kriging (see Wackernagel 2003 for a complete description of the theory of spatial filters). In a few words, filtering kriging decompose the spatial densities into *n*+1 component, where *n* is the number of spatial scales. The first component corresponds to the "nugget" of the variogram, that is the non-spatial component of the variability existing in the data. Spatial predictions associated to this first component result in a constant that i**s** the expected animal density calculated on the basis of the non-spatial variability of the data. In ecological term, this corresponds to a "basal" density of animals over the study area. In addition from this "basal" density, the spatial filters are gaussian, spatially structured random-fields that describe how the spatial variability modulates the basal density, and only contains spatial effects. There is one spatial filter per spatial structure modelled with the variogram. Variogram modelling and filtering kriging was achieved with hand-made R routines designed by Edwige Bellier. See Bellier (2007) for further details.

The oceanographic covariates:

The set of covariate was extracted daily and then averaged for 15 days time periods, from the OCEANWATCH database (http://las.pfeg.noaa.gov/oceanWatch/oceanwatch_safari.php). The set of covariate was constituted of data on Sea Surface Temperature (SST); Sea Surface

Height anomaly (SSHa); Surface Wind Strength (SWS); Surface Wind Divergence (SWD); Surface Chlorophyll a (CHLa) and Bathymetry (BAT). The associated gradients were computed using hand-made routines under R software. The spatial structure of each oceanic covariate was also examined with spatial correlograms.

Modeling FW spatial distribution:

Once the spatial structure existing in FW dataset and in environmental covariates identified, the expected "basal" density of FW (corresponding to the variogram nugget) has been calculated, and *n* spatial filters were extracted, *n* being the number of spatial structure identified in the FW dataset. Then, *n* spatial models were built, using Generalised Additive Models (Wood & Augustin 2001). These models assumed a Gaussian distribution for the spatial filters, a sustainable hypothesis since spatial filters issued from filtering kriging are gaussian random fields. According to the results of spatial correlograms, environmental covariates were also splitted in *n* groups according to their spatial structure. For each spatial model, all covariates presenting a spatial structure at a similar scale than the scale of the modelled spatial filter were tested, and the covariates used in each final spatial model were chosen according to a forward selection procedure. The final spatial models were then used to predict FW distribution at their respective scales, and these scale-dependent predictions were combined (summed) together and to the expected basal FW density in order to highlights the areas the most suitable for FW (i.e. potential distribution areas) at various dates (from June to August, for 15-days time periods).

RESULTS :

Pre-processing the FW dataset:

The uncorrected, segmented data set is shown in fig 1. Because FW observations were near-exclusively carried out in the southern part of the study area, we focused on it for further statistical analysis (Fig 1). The comparison between corrected and uncorrected data (Fig 2) showed that the number of FW was locally multiplied by a factor 3, with in some cases up to 25 FW predicted in the same segment (that is a density of near 5 FW per kilometre). Moreover, the linear relationships between corrected and uncorrected data was strongly significant ($p<2e^{-16}$ R^2=0.87). In addition, spatial variograms of uncorrected and corrected segmented data displayed quite different pattern (Fig. 3), with variogram on uncorrected data showing a small-scale spatial structure at 0.5 degree, while patterns on the variogram issued from corrected data showing chaotic variations. In these conditions, we considered that spatial models to identify the environmental parameters affecting FW distribution should be built using uncorrected data, because data correction lead to irrelevant local abundance of FW and moreover strongly deteriorates the spatial structure existing in the dataset. In addition, since uncorrected data correlates very well to corrected ones (Fig. 2), there is few chances that visibility bias induces spatial heterogeneities in the data that could lead to mis-identification of species-environment relationships with uncorrected data. By contrast, the artefacts issued from correction (i.e. artificially increasing FW local densities) may prevent from identifying species-environment relationships when using corrected data.

Spatial Structure of FW distribution:

The spatial correlogram (fig 4) carried out on the FW uncorrected dataset shows a strong positive spatial autocorrelation in the first distance class (up to 0.5 decimal degree), and a more diffuse, but always significant, positive spatial autocorrelation up to 2.5 decimal degree. The observed negative spatial autocorrelation at a scale of 4-8 decimal degree probably results from the large absence area in the north (see fig 1) and is not relevant for our study. The result obtained from the correlogram lead us to consider two spatial scales for the spatial distribution of FW, one small scale spatial structure at 0.5 decimal degree (near 30km) and another between 0.5 and 2.5 decimal degree (near 150 km). Therefore, we achieved a spatial decomposition of the FW dataset at these two scales with filtering kriging. The nugget component of the filtering kriging (i.e. the expected non spatial basal relative density of FW in the area) was equal to 0.33 individuals per segment. In addition, the Fig. 5 show the spatial patterns issued from the large and fine scale spatial filters, together with the data.

Spatial structure of oceanographic covariates:

The characterisation of the spatial structure of the oceanic covariates has been achieved with spatial correlograms and the results are reported in the table 1. Since FW distribution showed 2 spatial structure (one at 0.5 decimal degree, and the other from 0.5 to 2.5 decimal degree), our spatial covariates were splitted into two groups (table 1) in order to built two spatial models, one for each spatial filter. Covariates tested for inclusion in the spatial model at large scale are : bathymetry (TOPO), sea surface chlorophyll a (CHLA) and its associated gradient (CHLAg), sea surface temperature (SST), sea surface height anomaly (SSHg), and wind vector modulus (~wind strength, WINM). Covariates tested for inclusion in the small scale spatial model are bathymetry gradient (TOPOg), sea surface temperature gradient (SSTg), sea surface height anomaly gradient (SSHg), wind divergence (index of eckman pumping, WIND) and its associated gradient (WINDg), and wind vector modulus gradient (WINMg).

Scale dependent modelling of FW distribution:

Each spatial filter has been modelled with a set of environmental covariates presenting an equivalent spatial structure. The final set of covariates has been chosen following a forward-selection procedure and is based on the minimisation of the score of the gcv (Generalised Cross Validation) criterion. Each covariate received a penalty of maximum 5 knots to avoid data overfitting. Covariates retained for the large-scale model were WINM, CHLA, TOPO, SST and SSH (see Fig. 6). The large-scale model explained 42% of the deviance existing in the large-scale filter of FW spatial distribution and performed quite well (see fig. 7 for the diagnostic plots of the large scale model). Covariates retained for the small-scale model were WINDg, SSHg and SSTg, (Fig 8.) that is three spatial gradients issued from dynamic, environmental covariates. This fine-scale model only explained 8.31% of the deviance existing in the data, but again diagnostic plots (Fig. 9) showed that the model performed fairly well, even if probably less explanative than the large-scale one.

Scale dependent prediction of FW distribution:

The large and fine-scale predicted distribution of FW in the Bay of Biscay is shown in Fig 9 and 10, respectively, and Fig 11 shows the global predictions that result from the combination of the expected basal density of FW and of both spatial models. These predictions are provided for 15-days time period that range from June 2006 to August 2006. These maps shows that both the large-and fine-scale distribution of FW is subject to temporal variability, but three areas at large scale are identified as presenting environmental conditions suitable for FW distribution. These areas are located in the south-western corner of the study area, in the northern, central part and in the western-central part. Predictions issued from the large-scale models are more difficult to characterise but suggest that fine-scale patterns may be a factor that can add an important complexity to FW spatial distribution. Nevertheless, the global prediction of FW spatial distribution (Fig 11) seems to be highly influenced by large-scale patterns.

DISCUSSION :

The presented modelling approach is innovative in the sense that it focuses only on potential ecological processes driving FW distribution, by looking exclusively at environmental covariates that truly characterises their oceanic environment. Among the oceanic factors that were identified at large scale, we can notice the strong positive relationships between FW and Wind strength (WINM) (fig. 6), the positive relationships between FW and SST with an optimum in sea surface temperature around 19°C; and their association with extreme values of SSH. We can hypothesize that strong winds are a good index of surface water mixing that may enhance productivity in the water-column, that the 19°C optimum may be related to an optimum for the growth, reproductive success or survival success to their main preys; and that extreme SSH values indicates retention areas where resources may be aggregated. At small scale however, the identified relationships between FW and oceanic covariates were less convincing, probably because FW distribution at small scale is more likely to be influenced by other biological covariates that are closer to FW in the pelagic food web, such as zooplankton or small schooling fishes. Therefore, it is important to separate both large and small scale effects in the modelling process for management purpose. Indeed, the inference that can be drawn from our large-scale maps is more robust for FW than the inference that can be drawn from the small-scale one, given the performance of our large- and small- scale models. We then recommend, for management purpose, that proposal for example of protected areas or of area of controlled use should be based on the results provided by our large-scale maps only. Moreover, we should mention that the proposed modelling exercise is based on one survey only, carried out at a punctual date. Therefore, the situation described by the models only concerns the studied time period and all predictions obtained outside this time periods should be interpreted as “potential areas” for FW given that their relationships with the oceanic covariates remains the same than those measured during the survey. It is clear that such modelling exercise would greatly benefit from further surveys carried out at different seasons, in order to take the seasonal variability existing in the FW population into account during the modelling exercise.

REFERENCES:

Bellier, E. 2007. Spatial distribution within hierarchical patch dynamics systems: identifying species-environment spatial relationships at multiscales. Phd thesis, University of Mediterranean Sea, Marseille.

Bjørnstad, O. N. and Falck, W. 2001. Spatial covariance functions: estimation and testing. *Environmental and Ecological Statistics* 8: 53-70.

Bjørnstad, O. N. (2006). Spatial nonparametric covariance functions. R package available at http://asi23.ent.psu.edu/onb1/

Certain, G., E. Bellier, B. Planque, and V. Bretagnolle. 2007. Characterising the temporal variability of the spatial distribution of animals: an application to seabirds at sea. *Ecography* 30:695-708.

Cressie, N. A. C. 1993. Statistics for spatial data. Wiley inter-science, New York.

Fauchald, P., K. E. Erikstad, and H. Skarsfjord. 2000. Scale-dependent predator-prey interactions: the hierarchical spatial distribution of seabirds and prey. *Ecology* 81:773-783.

Kotliar, N. B. and J. A. Wiens. 1990. Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. *Oikos* 59:253-260.

Russell, R. W., G. L. Hunt Jr, O. K. Coyle, and R. T. Cooney. 1992. Foraging in a fractal environment: Spatial patterns in a marine predator-prey system. *Landscape Ecology* 7:195-209.

Pebesma, E. J. and Wesseling, C. G. 1998. Gstat: a program for geostatistical modelling,

prediction and simulation. *Computers and Geosciences* 24, 17–31.

Schneider, D. C. and J. F. Piatt. 1986. Scale-dependent correlation of seabirds with schooling fish in a coastal ecosystem. *Marine Ecology Progress Series* 32:237-246.

Wackernagel,H. 2003. Multivariate Geostatistics. Springer.

Wood, S. N. and N. H. Augustin. 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling* 157:157-177.

Fig 1. CODA study area, with bathymetry (coloured levels) and FW observations (purple squares). The polygon delineates the area on which spatial analysis were carried out.

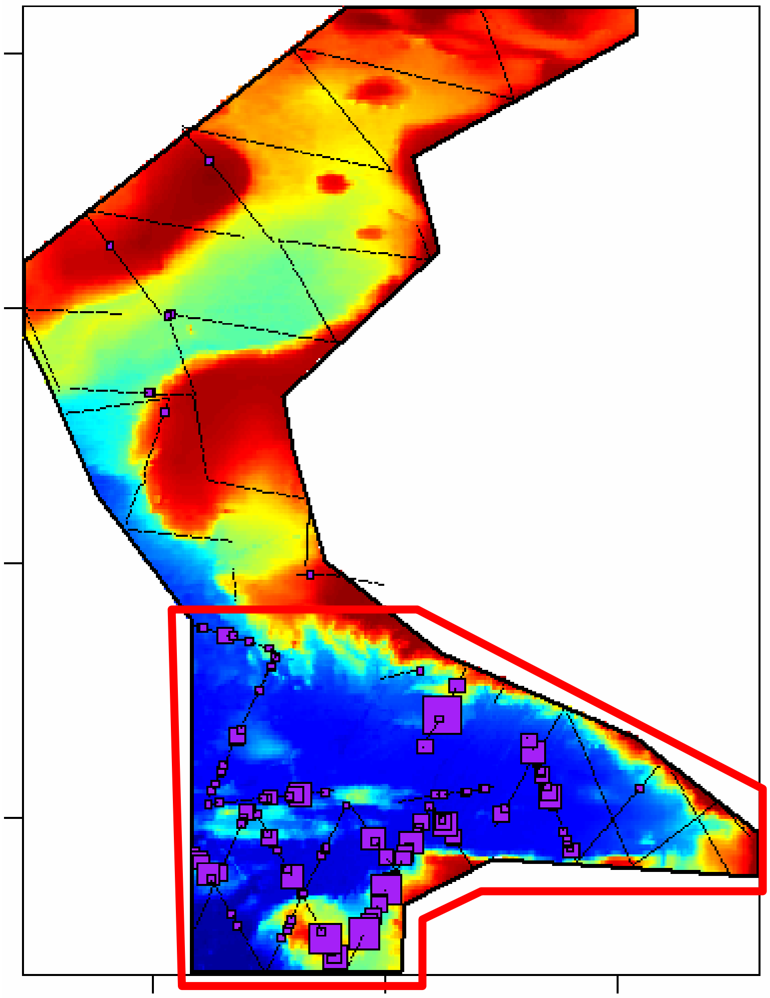

Fig 2. statistical relationships between uncorrected segmented data and segmented data corrected for visibility bias.

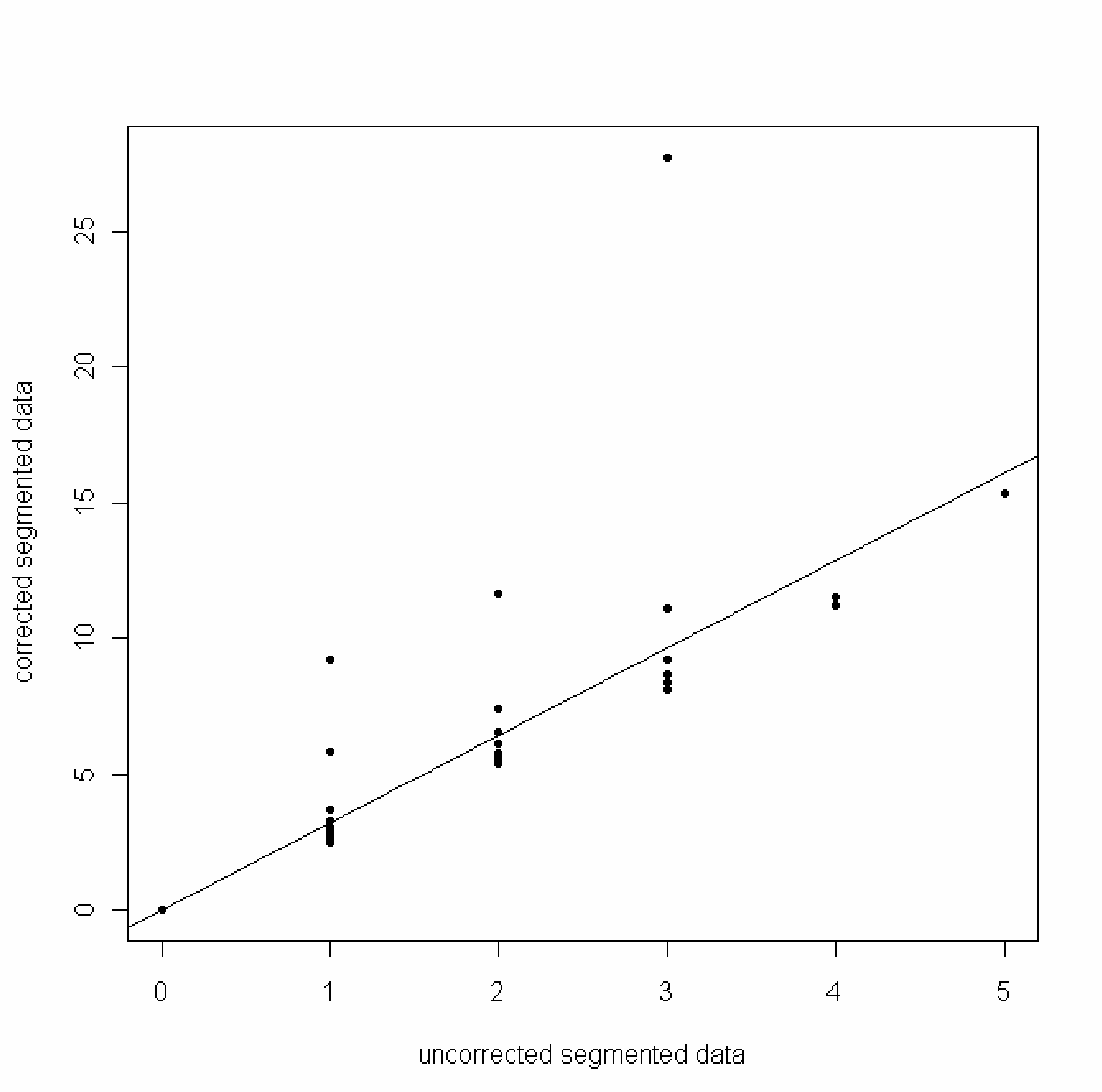

Fig 3. Experimental variograms showing the spatial structure of the variance existing in uncorrected segmented data and in segmented data corrected for visibility bias.

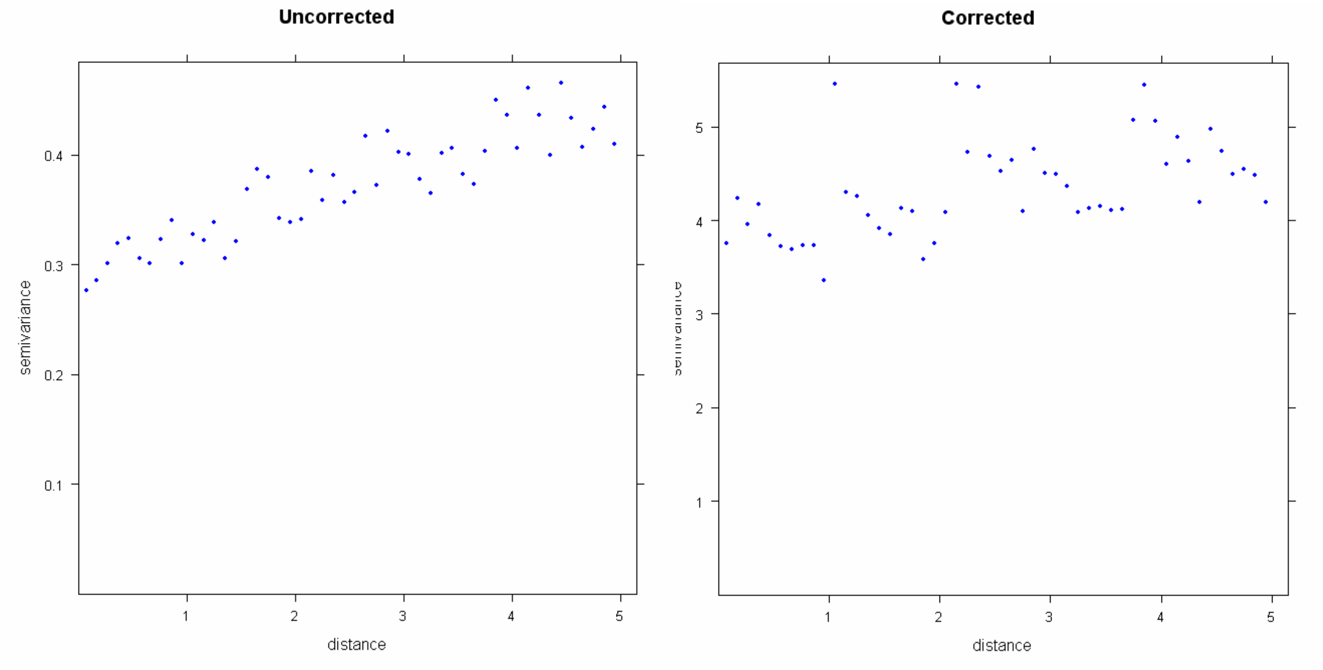

Fig 4. Correlogram of uncorrected, segmented data, and Variogram models fitted to the experimental variogram of FW data. The full variogram model (full line) is a nested model composed of a small-scale model (segmented lines) and a large-scale model (dotted lines).

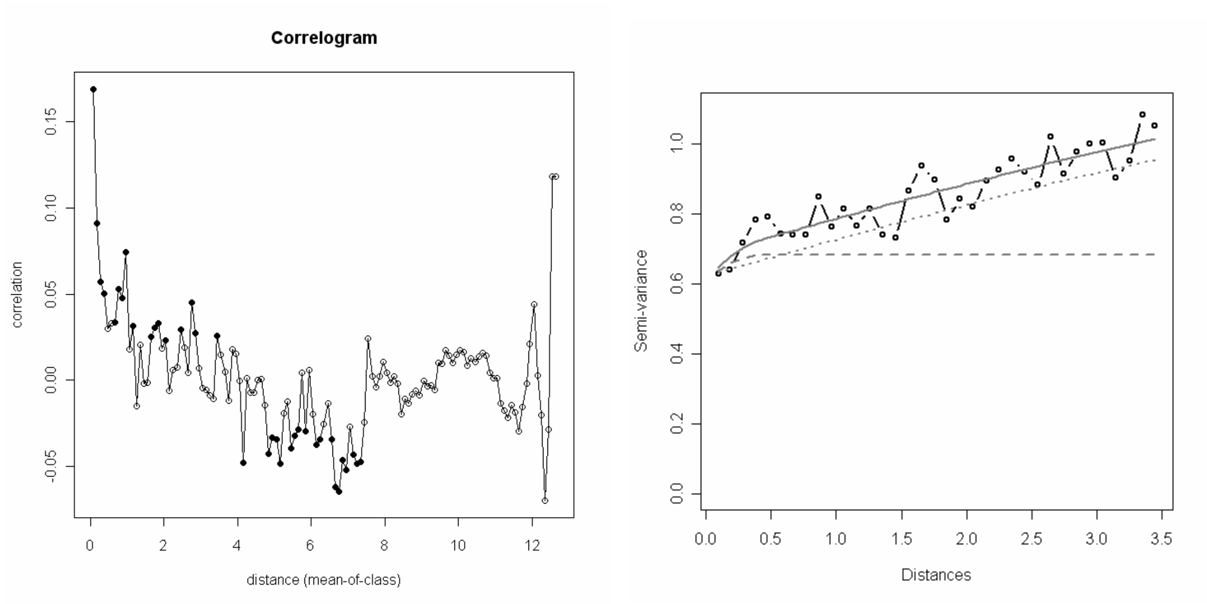

Fig 5. Spatial filters (red symbols) extracted at large and small scale, together with the segmented raw data (blue symbols)

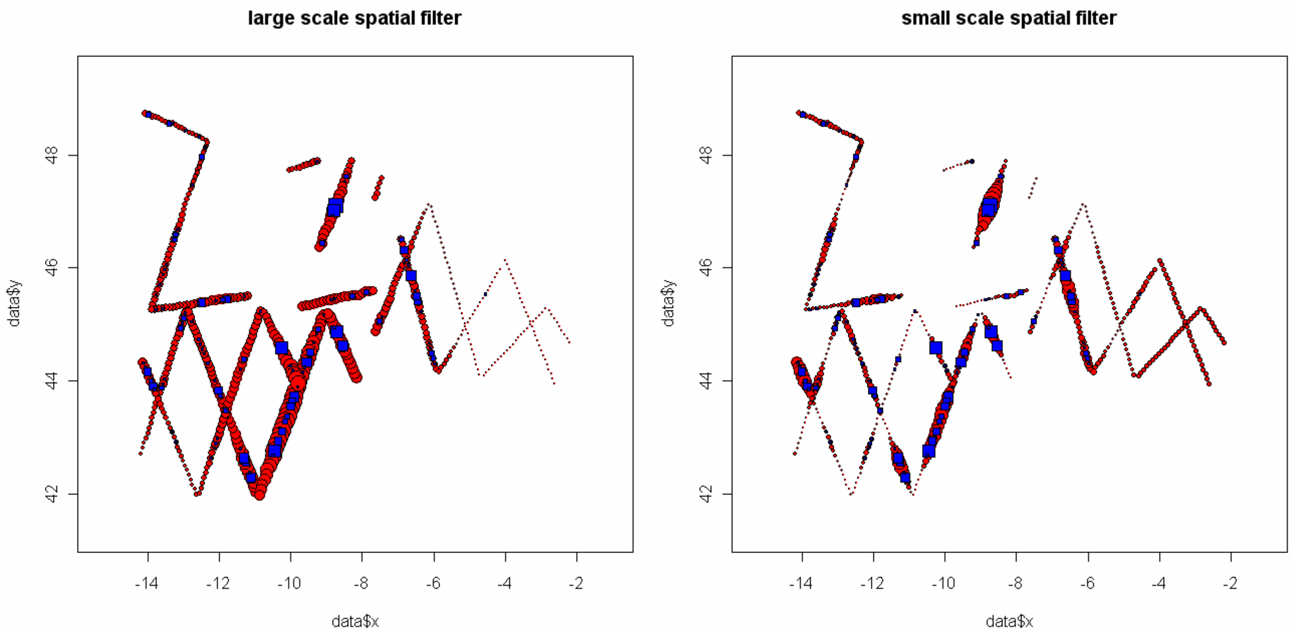

Fig 6. Relationships modelled between the large-scale filter of FW distribution and the environmental covariates.

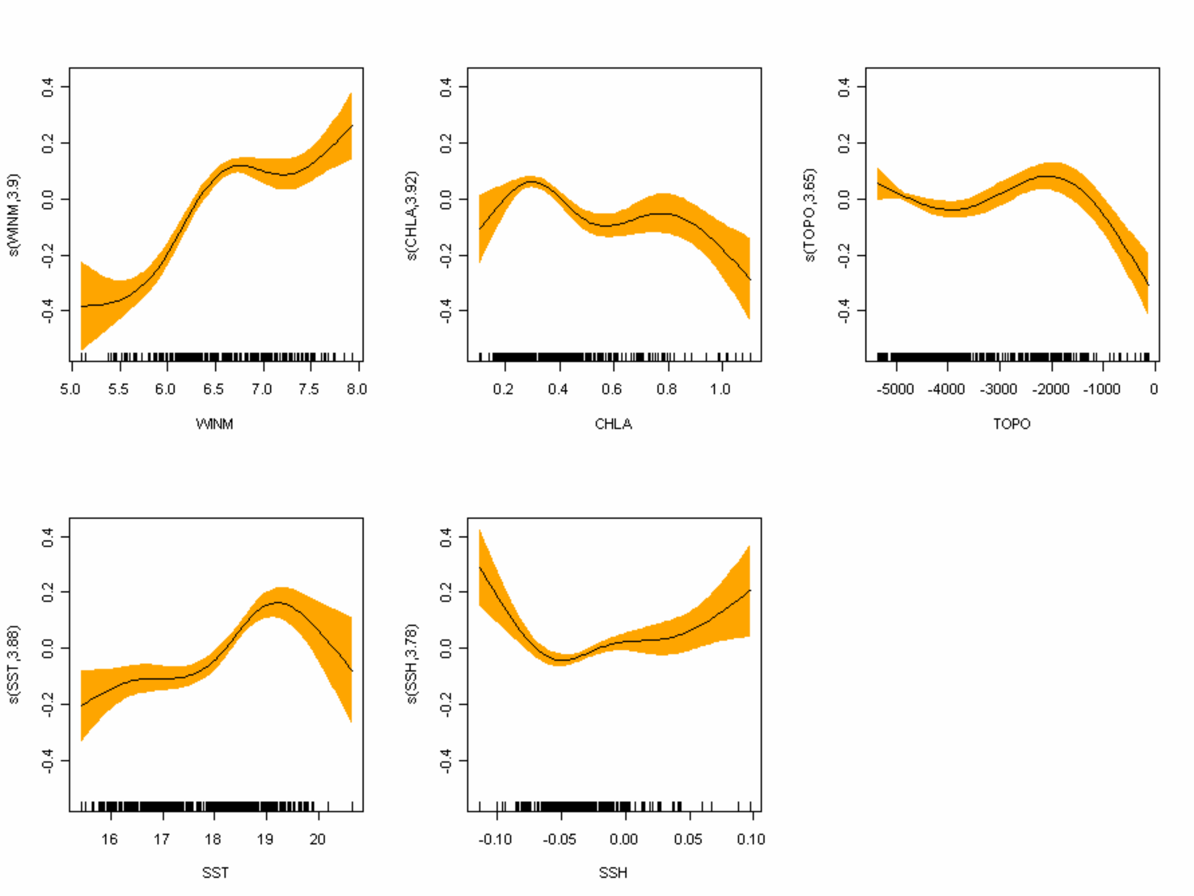

Fig 7. Diagnostic plots of the large-scale model

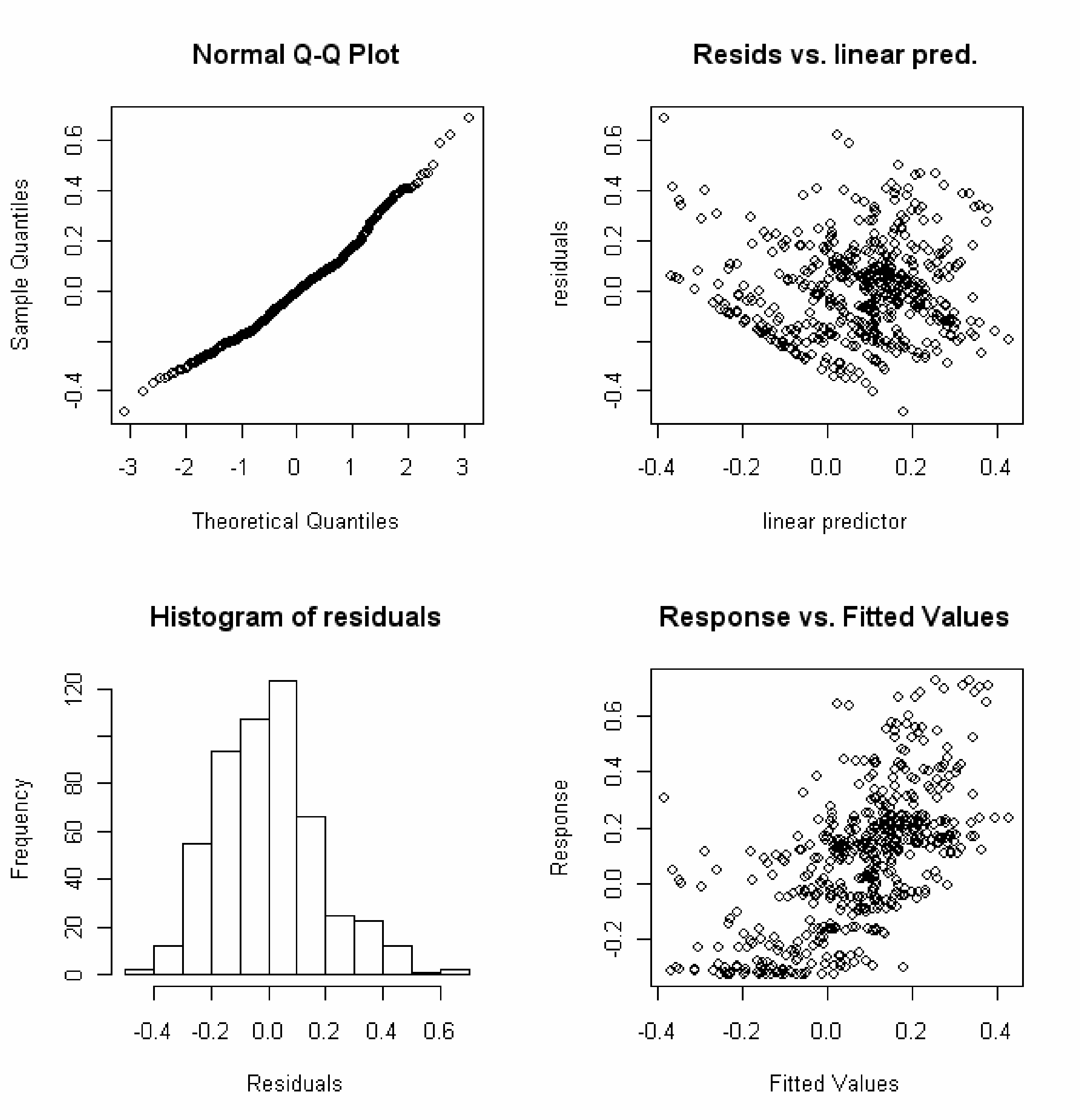

Fig 8. Relationships modelled between the fine-scale filter of FW distribution and the environmental covariates

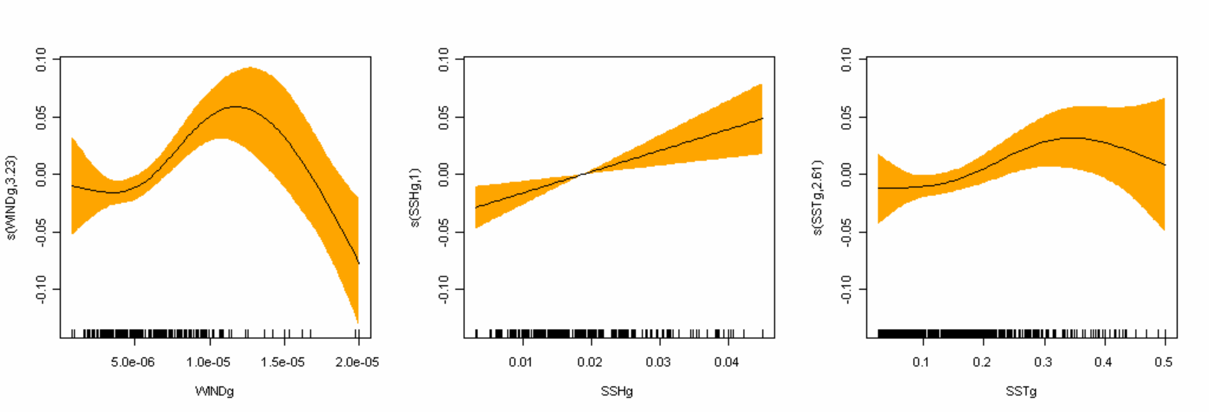

Fig 9. Diagnostic plots of the fine-scale model

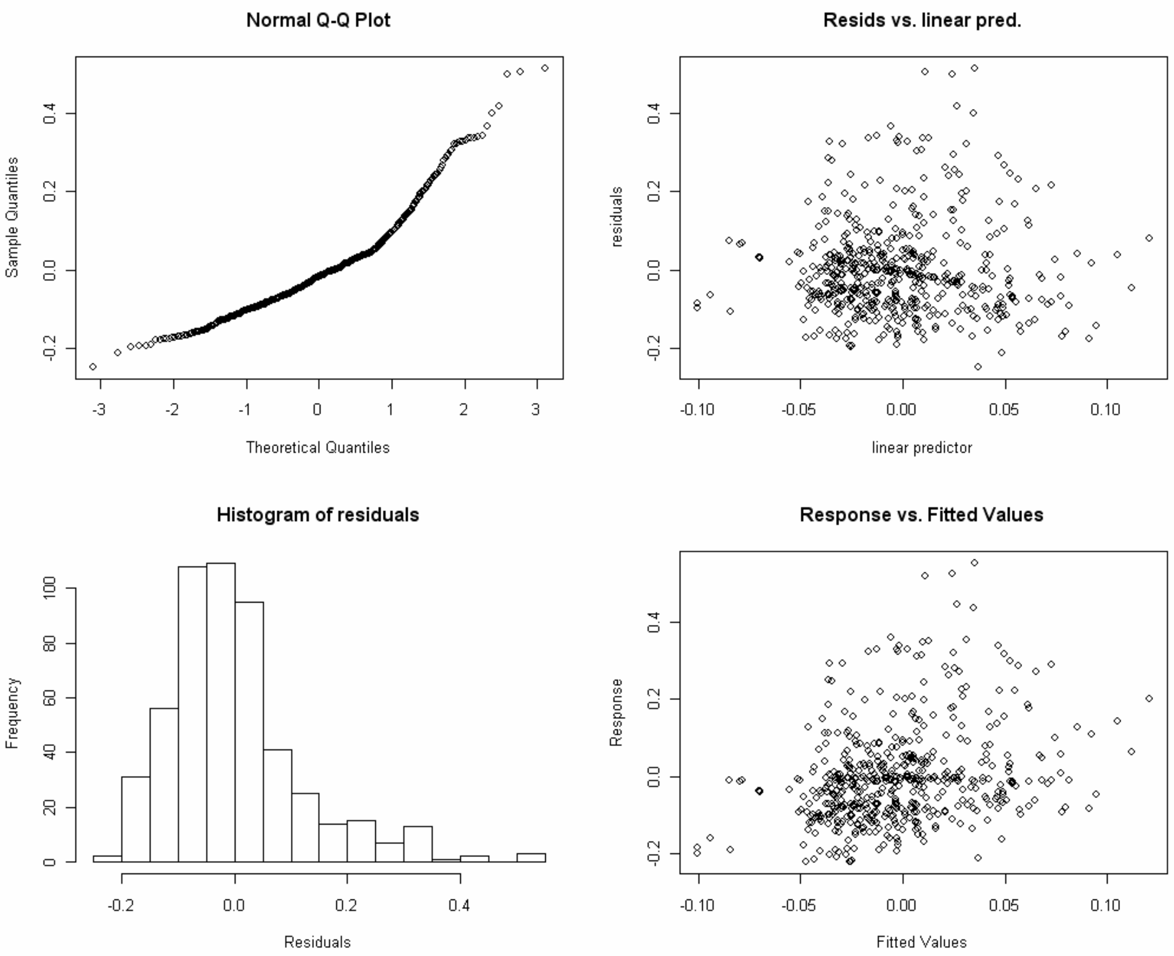

Fig 10. Spatial prediction of FW distribution issued from the large-scale model, for different time periods of 15 days. Black squares show the data.

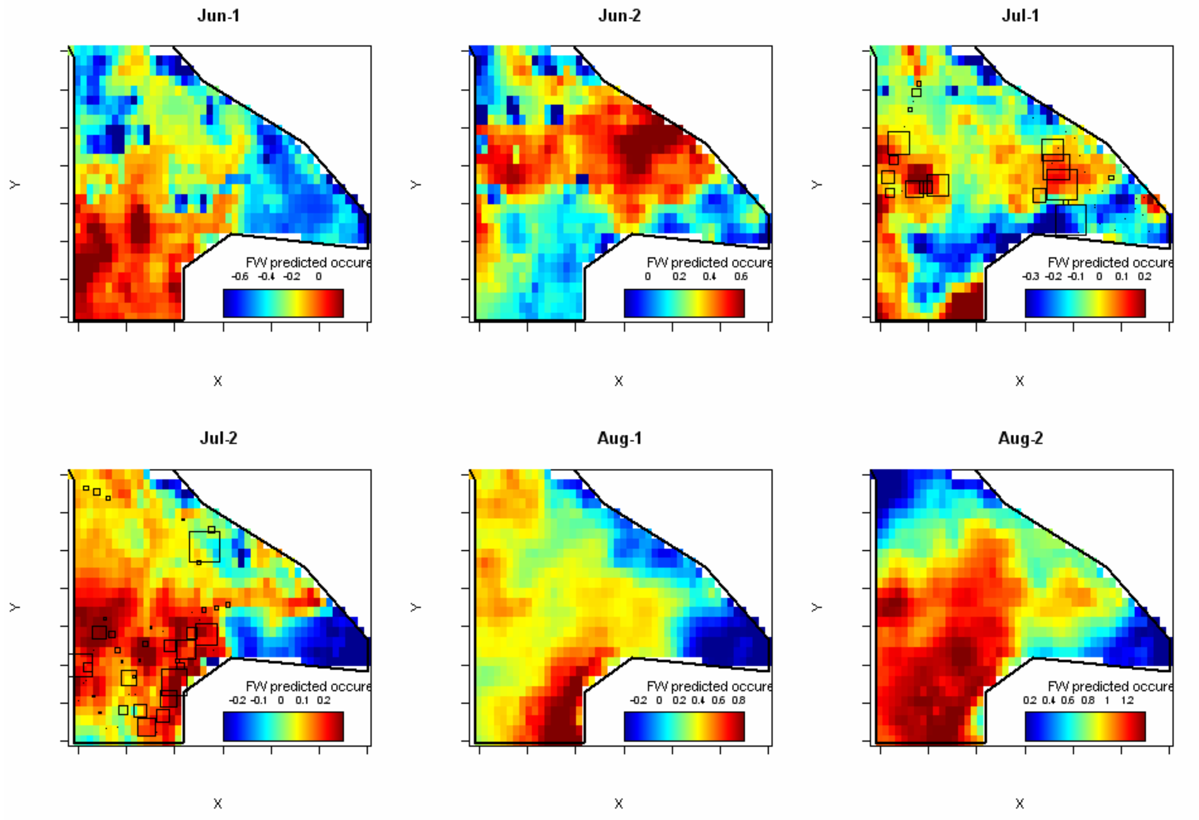

Fig 11. Spatial prediction of FW distribution issued from the small-scale model, for different time periods of 15 days. Black squares show the data.

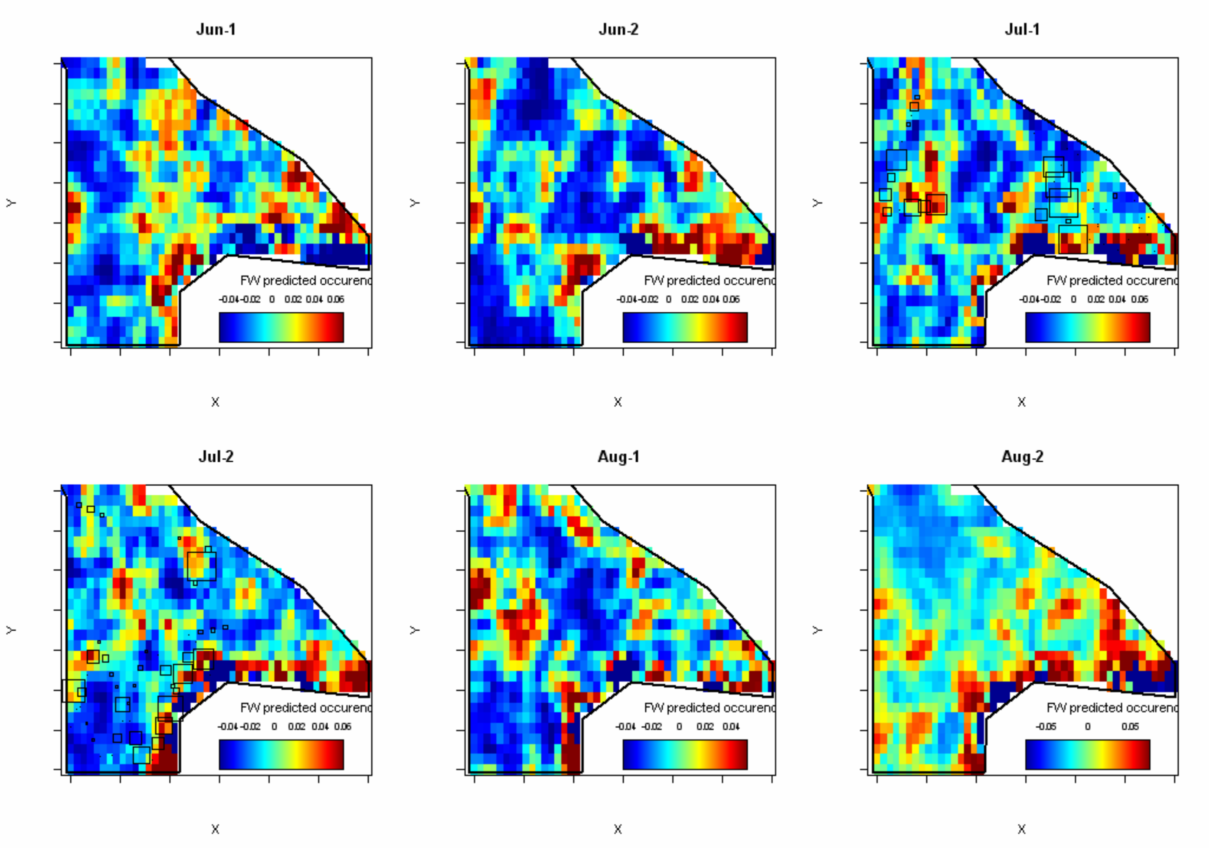

Fig 12. Final predictions maps of FW relative density, obtained by summing the expected basal density issued from the non-spatial variability of FW data, and the spatial effects issued from spatial models at large and fine scale. Black squares show the data.

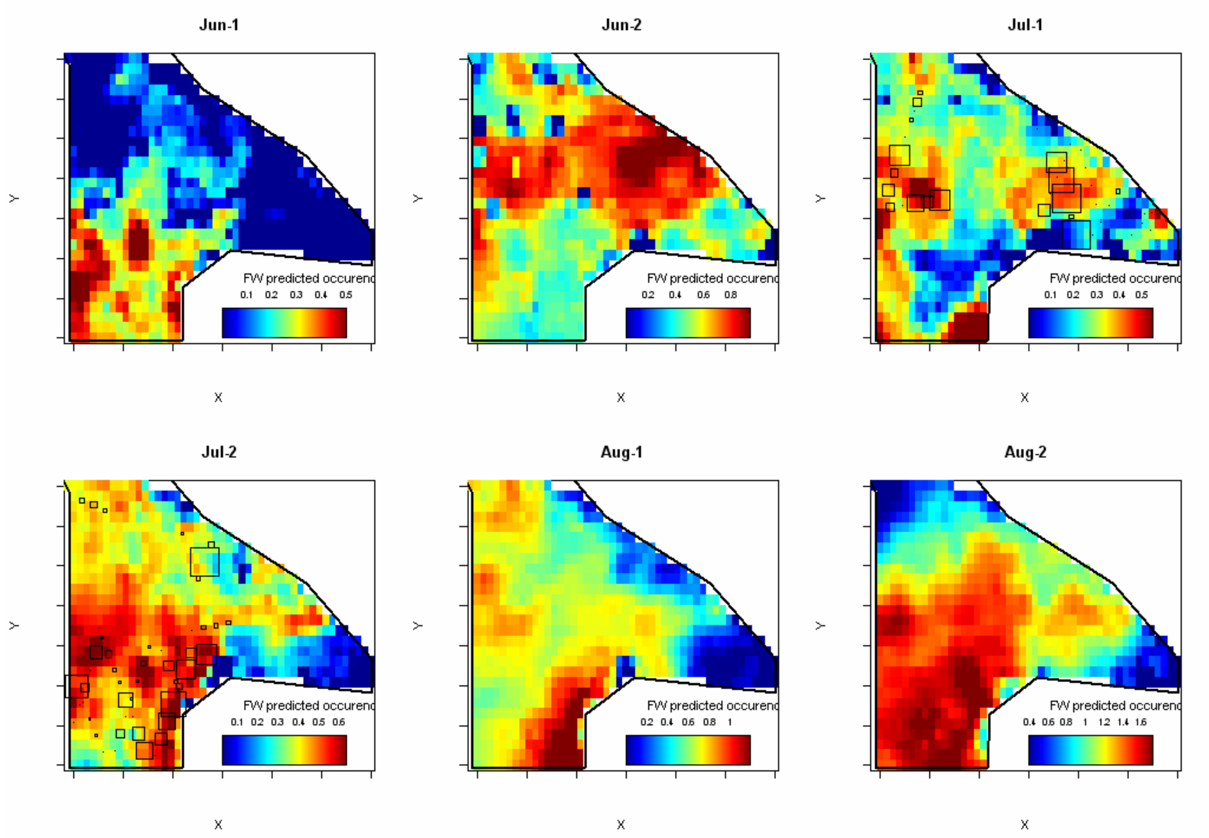

Table 1. Summary of the spatial structure of the environmental covariates, in decimal degree.

Covariate	scale	to be related to…
TOPO	2	large scale filter
TOPOg	0.4	small scale filter
CHLA	3	large scale filter
CHLAg	2	large scale filter
SST	5	large scale filter
SSTg	0.5	small scale filter
SSH	2	large scale filter
SSHg	0.9	small scale filter
WIND	0.9	small scale filter
WINDg	0.7	small scale filter
WINM	2.5	large scale filter
WINMg	0.4	small scale filter