European Union's Seventh Framework Programme

Grant Agreement **Nº: 603521**

Project Acronym: **PREFACE**

Project full title: **Enhancing prediction of tropical Atlantic climate and its impacts**

Instrument: Collaborative Project

Theme: ENV.2013.6.1-1 – *Climate-related ocean processes and combined impacts of multiple stressors on the marine environment*

Start date of project: 1 November 2013 Duration: 48 Months

Deliverable reference number and full title: D 10.2: " Bayesian hierarchical model: Report on the Bayesian hierarchical model developed in Task 10.1, including an assessment of previous hierarchical Bayesian approach-based work in the field of climatology, Software of the models developed in Task 10.4."

Lead beneficiary for this deliverable: 11

Due date of deliverable: 31.10.2016

Actual submission date: 31.10.2016

Project co-funded by the European Commission within the Seven Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the Consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

If this report is not to be made public, please state here why: Part of results described below are included in a manuscript currently under review for publication by Nature Communications. As such, the authors prefer to not divulge the entire contents of this deliverable to the wider public until further notification from Nature Communications.

Contribution to project objectives – with this deliverable, the project has contributed to the achievement of the following objectives (from Annex I / DOW, Section B1.1.):

N.º	Objective	Yes	No
1	Reduce uncertainties in our knowledge of the functioning of Tropical Atlantic (TA) climate, particularly climate-related ocean processes (including stratification) and dynamics, coupled ocean, atmosphere, and land interactions; and internal and externally forced climate variability.	X	
2	Better understand the impact of model systematic error and its reduction on seasonal-to-decadal climate predictions and on climate change projections.	X	
3	Improve the simulation and prediction TA climate on seasonal and longer time scales, and contribute to better quantification of climate change impacts in the region.		
4	Improve understanding of the cumulative effects of the multiple stressors of climate variability, greenhouse-gas induced climate change (including warming and deoxygenation), and fisheries on marine ecosystems, functional diversity, and ecosystem services (e.g., fisheries) in the TA.		
5	Assess the socio-economic vulnerabilities and evaluate the resilience of the welfare of West African fishing communities to climate-driven ecosystem shifts and global markets.		

Author(s) of this deliverable: Angelo Rubino, Carlo Gaetan, Davide Zanchettin, Maeregu Woldeyes Arisido, Jorge Lopez Parages

Deviation from planned efforts for this deliverable: none to our knowledge.

Report

Executive Summary: WP10: "Statistical methods to assess and improve forecast of Tropical Atlantic variability" aims at (1) developing a Bayesian hierarchical modeling strategy to re-calibrate forecasts and improve prediction of Tropical Atlantic variability (TAV) and its impact, (2) developing a statistical scheme to predict sea surface temperature (SST) anomalies in remote regions associated with TAV, and (3) assessing the ability of state-of-the-art climate models (CMIP5) to reproduce climate variations over the Tropical Atlantic Sector (including surrounding continents) over the 20th century. To the first purpose, Bayesian hierarchical modeling strategies are being developed and used at the University of Venice to improve TAV predictions and re-calibrate simulated data for bringing them into line with measurements. The main theoretical advantage of a Bayesian strategy is that it encapsulates the uncertainties involved in the estimation of all model parameters and these uncertainties are properly taken into account in the predictions and re-calibrations. This deliverable

presents the status of the work progress concerning the development of the general Bayesian methodology to improve TAV prediction, identify sources of heterogeneity and non-stationarity of simulated TAV features, and to unify different spatial-temporal dependences, including fronts and shifts, into a unique statistical framework (Task 10.1). Specifically, two branches of the model's development have been pursued, which target different components of the issue: the first branch is focused on estimation of the temporal component of systematic model errors, using the evolution of SST drifts in the Tropical Atlantic region from decadal climate predictions as a test bed and case study; the second branch is focused on the spatial assessment of the bias in a multi-model ensemble, using near-surface air temperatures over the Tropical Atlantic region from CMIP5 historical simulations as test bed. The latter model also contributes to Task 10.2. This deliverable further includes R code used to deal with the spatial misalignment between observational and model data, or by the scarceness/incompleteness of observations (Task 10.4). The proposed strategy optimizes the one-way rescaling between numerical climate model outputs, provided on the original model grid, and grid/point observations.

The present document consists of three main sections:
1 – Review of scientific literature on hierarchical Bayesian approach-based work in the field of climatology
2 – Report on the Bayesian hierarchical model
3 – Description of statistical models for spatial misalignment

1 – Review of scientific literature on hierarchical Bayesian approach-based work in the field of climatology

Bayesian hierarchical models (BHMs) use Bayes' theorem to incorporate information from different sources, including observations, physical theories and experts' knowledge. They provide a flexible framework to developing consistent inference and prediction of unknown quantities under study, which overcomes single-value predictions in the assessment of uncertainty.

The idea of hierarchical modeling of scientific processes largely stems from the work of Berliner (1996), who, in a period of fast development and popularization of Markov Chain Monte Carlo (MCMC) methods in Bayesian statistics, proposed BHMs as a probabilistically consistent way to partition uncertainty in systems with complicated data, process and parameter relationships. Wikle et al. (2013) provide a recent review on the development of the hierarchical model paradigm in the field of natural sciences, with a special focus on oceanography, and illustrates how the Bayesian approach surpasses, in certain aspects, likelihood-based inferences. For instance, given the complexity of most ocean and atmospheric processes, the multivariate spatio-temporal dependence structures associated with the unobserved geophysical process of interest can be very complicated, including nonlinear temporal behavior and nonstationary in spatial and/or temporal features, and potentially very high-dimensional: In a likelihood-based approach the bulk of such possibly complicated dependence structures must be realistically specified in the statistical model; in a hierarchical approach, focus is on the unobserved process of interest directly, whose modeling can therefore directly incorporate insights about it, such as Markovian approximations of it (like in

dynamic linear models), or more complex mathematical representations of it (like simpler approximations of the laws of fluid-dynamics, or of spatio-temporal characteristics). A hierarchical approach further disentangles the measurement uncertainty from the process uncertainty and co-variability.

The BHM general formulation is based on three building blocks: the data model, the process model and the parameter model. Suppose that $\boldsymbol{Z}$ represents the data, $\boldsymbol{Y}$ represents the process and $\boldsymbol{\Theta}$ represents the parameters related to the data or the process model, we then have:

(a) Data model, $[\boldsymbol{Z}|\boldsymbol{Y},\boldsymbol{\Theta}]$

(b) Process model, $[\boldsymbol{Y}|\boldsymbol{\Theta}]$

(c) Parameter model, $[\boldsymbol{\Theta}]$

where [A] is the generic notation for the (conditional) probability distribution of the random quantity A. In practice, (a) defines the statistical model representing the dependence of observations on the unknown process, (b) describes the conditional probability distribution of the process on the model parameters, and (c) describes the (prior) probability distribution of the parameters, which are treated as random quantities according to the Bayesian approach.

BHMs provide estimates of the unknowns ($\boldsymbol{Y}$ and $\boldsymbol{\Theta}$) with associated uncertainty through the calculation of their posterior distribution conditioned to available observations, which is allowed by the Bayes' theorem:

$$[\boldsymbol{Y},\boldsymbol{\Theta}|\boldsymbol{Z}] = [\boldsymbol{Z}|\boldsymbol{Y}]\,[\boldsymbol{Y}|\boldsymbol{\Theta}]\,[\boldsymbol{\Theta}] \,/\, [\boldsymbol{Z}] \qquad (1.1)$$

The major difficulty in evaluating (1.1) relies on the computation of the predictive distribution $[\boldsymbol{Z}]$ that can be circumvented by means of a MCMC method (Robert and Casella, 2004).

Scientific literature entails numerous studies describing the application of BHMs in the field of climate science, especially to the assessment of climate model outputs and to paleoclimate reconstructions. We propose in the following a critical review of selected contributions, focusing on those that are most relevant for PREFACE WP10. Additional reviews of scientific literature about statistical models to combine climate model output can be found in Tebaldi and Knutti (2007), and Smith et al. (2009). Giorgi and Mearns (2002), Christensen et al. (2010), Knutti et al. (2010), and Knutti (2010) provide additional methodological approaches and discussions on the topic.

Several studies have contributed to establish the Baysian framework as attractive for combining information from different climate models (so-called multi-model ensembles): It decomposes the complicated relationship between the observations and the outputs of different models into simpler, hierarchical relationships that can be described in a reasonable and transparent way. Scientific work proposing statistical approaches for combining climate model output in the context of a Bayesian hierarchical model include both studies assuming independence among ensemble members (e.g., Tebaldi et al., 2005; Furrer et al., 2007; Smith et al., 2009; Tebaldi and Sanso, 2009; Kang et al., 2012) or explicitly accounting for model-to-model correlation (Sain and Furrer, 2009; Sang et al., 2011).

Tebaldi et al. (2005) were among the first to use the Bayesian framework for analysing multi-model ensembles. They proposed a Bayesian statistical model to combine information from a multi-model ensemble of coupled climate model simulations and observations to determine probability distributions of future temperature change on a regional scale. The aim was to produce probabilistic forecast of climate change that surpassed the reliability ensemble averaging method for multi-model evaluation that consists, essentially, of a weighted average of the individual model output, where the weights measure the criteria of model bias (with respect to observed climate) and convergence (with respect to projected climate change) implicitly assigned to the ensemble members. The study thus provided the first step from a heuristic criterion to a formal statistical model for climate model evaluation. Tebaldi et al. (2005) adopted uninformative prior distributions for their study focused on regional temperatures, and demonstrated that their model produced meaningful posterior distributions: The probability distributions of temperature change revealed features such as multimodality and long tails that could not otherwise be easily discerned.

The work by Tebaldi et al. (2005) has been generalized in several directions. Buser et al. (2009) extended the work by Tebaldi et al. (2005) by considering not only the long-term climate mean, but also the interannual variations, further accounting for possible nonstationarity of the data in the control and scenario periods. Their Bayesian model was also used to investigate the impacts of two different assumptions for extrapolating bias from the control period to the scenario simulations. First, they assumed that bias changes are negligible compared to climate change, or equivalently that the bias itself is time-invariant, therefore calling this assumption "constant bias" (see also the critical review of this assumption by Maraun, 2012). Then, they alternatively assumed that biases depend (linearly) on the underlying climate and thus change with time, calling this assumption "constant relation". Furthermore, based on the substantial biases they identified in interannual variability for various regional climate models they also argued that the same models would inaccurately estimate a future warming. The model suffered from intrinsic identifiability problem, as the data did not allow a clear separation between bias changes and climate changes, thereby requiring additional assumptions: specifically the identifiability problem was resolved by using informative priors for the bias changes. The results of an application to Alpine seasonal temperatures demonstrated that different, plausible assumptions about the bias (constant bias versus constant relation) can lead to substantially different estimates of future values for the climatic variable of interest. The work by Buser et al. (2009) have been further extended in several directions, for instance Ho et al. (2012) discuss two bias assumptions and their influence on future projections of European temperature.

Furrer et al. (2007) propose a BHM to combine current climate observations and simulations, as well as future climate projections, defining climate change as a spatial process on the sphere. Their model separates the spatial response into a large scale climate change signal and an isotropic process representing small scale variability across climate models. The formulation of their model illustrates the general methodology and basic spatial elements of the problem of assessing ensemble climate model outputs on spatial fields, which is relevant for the BHM developed within PREFACE WP10. We therefore briefly summarize it here.

Climate change in model i =1, …,N within the ensemble, is defined, in the different grid points of the model domain, as $\mathbf{D}_i = \mathbf{Y}_i - \mathbf{X}_i$, where $\mathbf{X}_i$ and $\mathbf{Y}_i$ are current and future scenario average climate,

respectively. The BHM is then built on the statistical regression model: $\mathbf{D}_i = \mathbf{M}\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i$, where M is a matrix of spatial basis functions, $\boldsymbol{\theta}_i$ is a vector of coefficients describing a random effect that is different for each model but whose expected value is the true difference in climate, and $\boldsymbol{\varepsilon}_i$ is a mean zero spatial process. The flexibility of this spatial random effects model allows for more general applications, such as that to assessing climate model biases as done within PREFACE WP10. At the *data level*, the BHM by Furrer et al. (2007) models $\mathbf{D}_i$ as a spatial process, specifically based on the assumption that the climate change is an additive decomposition of a large scale climate signal and small scale signals consisting of model bias and internal model variability:

$$\mathbf{D}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i,\ [\mathbf{D}_i \mid \boldsymbol{\mu}_i, \varphi_i\ ] \sim(\text{iid})\ N_n(\boldsymbol{\mu}_i, \varphi_i\boldsymbol{\Sigma}),\ \varphi_i > 0,\ i=1,...,N \qquad (1.2)$$

where | means "conditioned on", N_n indicates an n-dimensional normal density, $\boldsymbol{\Sigma}$ is the correlation matrix, and φ_i are scale parameters. The formulation above states that the density of the entire dataset, conditional on $\boldsymbol{\mu}_i$ and φ_i, is the product of N Gaussian density functions dictated by $N_n(\boldsymbol{\mu}_i, \varphi_i\boldsymbol{\Sigma})$. A similar separation is performed in the BHM developed within PREFACE WP10.

At the *process level*, the BHM reduces the high-dimensional large-scale signal by imposing $\boldsymbol{\mu}_i = \mathbf{M}\boldsymbol{\theta}_i$, where $\mathbf{M}$ contains p<<N basis functions. Further, assuming that climate models depart from the true large-scale climate change pattern $\boldsymbol{\vartheta}$ with covariance $\Psi_i\boldsymbol{\Omega}$, and that the ensemble overall does not exhibit systematic large scale errors, it is posed:

$$[\boldsymbol{\theta}_i \mid \boldsymbol{\vartheta}, \Psi_i\ ] \sim(\text{iid})\ N_p(\boldsymbol{\vartheta}, \Psi_i\boldsymbol{\Omega}),\ \Psi_i > 0,\ i=1,...,N \qquad (1.3)$$

where Ψ_i captures the different levels of bias and internal variability across the ensemble members. A similar decomposition is performed in the BHM developed within PREFACE WP10. At the *parameters level*, priors are defined for the unknowns, namely Ψ_i, φ_i, and $\boldsymbol{\vartheta}$. The work by Furrer et al. (2007) appears relevant also for their assessment of different choices about the basis functions, including spherical harmonics and indicator basis functions, and about the assumption of isotropy and stationarity of the error spatial process $\boldsymbol{\varepsilon}_i$. The number of basis functions is the most critical parameter identified by Furrer et al. (2007): a richer basis will capture more small scale features as $\boldsymbol{\vartheta}$, while leaving less for the small scale component $\boldsymbol{\varepsilon}_i$. The authors also demonstrated that the sensitivity to the number of basis functions increases for regional analyses on a subset of the spatial domain. Similar criticalities have been found and addressed in PREFACE WP10 (see section 2.4).

Jun et al. (2008) employed a nonstationary spatial process model to evaluate the bias of each climate model within a 20-member multi-model ensemble with contributing to the fourth assessment report of the Intergovernmental Panel on Climate Change, and used kernel smoothing to estimate the correlations of biases across the different climate models. They found that most of the climate model bias patterns are correlated. In particular, climate models developed by the same institution have highly correlated biases. They also found evidence that the model skills for simulating the mean climate and simulating the warming trends are not strongly related. Despite Jun et al. (2008) do not follow a Bayesian approach (they use instead maximum likelihood estimates), their study is relevant here as it surpassed the assumptions that each climate model is independent from the others and is a random sample from a distribution with the true climate as its mean. This so-called truth-centered approach implies that the average of a set of models converges to the true climate as more and more

models are added. The model bias D is defined as a process in space and time, and decomposed as $D_i(\mathbf{s},t) = b_i(\mathbf{s}) + u_i(\mathbf{s},t)$, where b_i is a purely spatial field with possibly nonstationary covariance structure and represents the bias of the i^{th} model with respect to the observed climate, and the residual u_i has zero mean and is assumed to be independent of b_i, and describes measurement errors as well as simulated climate variability. The spatial field b_i is modeled as a Gaussian random field with a mean structure depending on several covariates, including latitude (L), altitude (A) and a stochastic nonstationary zero-mean Gaussian process a_i:

$$b_i = \mu_{0i} + \mu_{1i}L(\mathbf{s}) + \mu_{2i}\mathbf{1}(\mathbf{s} \in land) + \mu_{3i}A(\mathbf{s}) + a_i(\mathbf{s}) \quad (1.4)$$

As a focus of their study, Jun et al. (2008) investigate the correlation between biases in different models, which poses the need to jointly model the cross-covariances $\sigma_{ij}(\mathbf{s}) = cov\{a_i(\mathbf{s}), a_j(\mathbf{s})\}$. To fill this need, a kernel smoother is applied to the statistics

$$\tilde{D}_{ij}(\mathbf{s}) = \sigma_{ij}(\mathbf{s}) + \varepsilon_{ij}(\mathbf{s}) \quad (1.5)$$

where $\varepsilon_{ij}(\mathbf{s})$ is a spatial process with zero mean. Then, a positive definite Gaussian kernel estimator is applied to $\sigma_{ij}(\mathbf{s})$.

Neeley et al. (2014) propose a spatial confirmatory factor analysis (FCA) model to combine the output of multiple climate models within an ensemble and characterize modes of similarity among the ensemble members. Despite not focused on model bias, the model of Neeley et al. (2014) is relevant in this review as it uses both Bayesian and spatial methods to estimate a common climate factor and model-specific, spatially correlated factor loadings for each ensemble member indicating the degree of agreement for each member with the common climate factor. Application of the statistical model to the ensemble of regional climate models participating to the North American Regional Climate Change Assessment Program demonstrated how the proposed approach allows identifying areas of disagreement between the different models, hence areas where no suitable consensus can be obtained.

In the FCA model by Neeley et al. (2014), climate is considered as a latent variable common to all models, i.e., which underlies the output of each climate model in the ensemble. FCA is different from an (optimal) weighted mean approach as those described above, as FCA extracts the features that the ensemble members have most in common (i.e., the consensus process) while down-weighting the features of divergent members, whereas the weighted average method gives the most weight to ensemble members with unique values. The identified factor process does not necessarily correspond to the actual climate process of interest; it represents instead an ensemble consensus view of the process.

The one-factor CFA model for grid-point i is written in Neeley et al. (20014) as:

$$y_{ij} = \lambda_{ij}\,\delta_i + \varepsilon_{ij} \quad (1.6)$$

where δ_i is the factor score in i (representing the consensus), λ_{ij} is the factor loading for the j^{th} model in i, and ε_{ij} is a Gaussian error $\sim N(0,\sigma^2_j)$. Neeley et al. (2014) express the spatial factor model in a

matrix form, and build it hierarchically with three levels, including likelihood (data), spatial structure of the factor loadings (process) and parameters.

Some methods, including that followed by Salazar et al. (2011), reduce the parameter space using eigenvalue decomposition, or a similar approach, to aid with computability.

The Bayesian framework has also been followed in a number of studies dealing with the forecast of a variety of geophysical systems (for a background on this application of BHMs, see Berliner et al., 2003). Milliff et al. (2011) present one of the latest steps in adapting BHM to large state-space systems with relatively abundant, multi-platform observations and process models motivated by geophysical fluid dynamics, and have demonstrated that the BHM methodology is viable in the operational ensemble ocean forecast setting. Specifically, Milliff et al. (2011) developed a BHM to estimate surface vector wind fields and associated uncertainties, to be used as forcing input data in the assimilation step of an ocean forecast system for the Mediterranean Sea. The operational application of a BHM strategy in the field of data assimilation for weather/ocean/climate forecasting systems must face the large volume of the employed datasets and high physical sophistication in the design of the model at the process level (see, e.g., compare the increasing complexity in Royle et al., 1998; Wickle et al., 2001; Milliff et al., 2011). Milliff et al. (2011) review the use of problem-specific assumptions driven by, e.g., data availability, process-model approximations and feasibility of implementation. The paper highlights how the BHM formalism renders explicit the assumptions regarding different levels of the BHM, such as the data level (e.g., the error properties of observations), the process level (e.g., the employed approximate physics) and the overall Bayesian solution procedures.

In brief, the BHM in Milliff et al. (2011) is built as follows: Let W be a vector of surface vector wind for all model grid locations and every input time within an assimilation cycle and let [W] be the probability distribution of the realizations of W. The properties of [W] can be used to quantify the best estimates of the surface vector wind and their uncertainties. The estimates and uncertainties in W are functions of relevant observations D and prior estimates or models for the surface vector wind. The posterior distribution for the surface vector winds, given data, is given by Bayes' theorem, following (1.1):

$$[W, \theta_d, \theta_p \mid D] \propto [D \mid W, \theta_d]\,[W \mid \theta_p]\,[\theta_p, \theta_d] \qquad (1.7)$$

where the three terms on the right are the model hierarchies to be described through the BHM: the data level $[D \mid W, \theta_d]$, the process level $[W \mid \theta_p]$, and the parameters level $[\theta_p, \theta_d]$. In this general formulation, measurement error models and/or retrieval algorithms are parameterized in $[\theta_d]$, while $[\theta_p]$ contains parameterizations of the process, necessary to determine the prior information.

The model of Milliff et al. (2011) further increases the complexity of the individual levels, for instance including an additional data-stage distribution upon that for the surface vector wind for sea-level pressure observations. The process model includes a geostrophic–ageostrophic partition of the surface vector wind with an explicit treatment of surface friction effects through the Rayleigh friction equations. The latter are implemented in the Bayesian model in a way that the zonal and meridional velocity components do not depend on each other, which is achieved through including sea-level

pressure spatial pattern terms. The stochastic and discrete form of the system equations in the process model is:

$$\mathbf{U}_t = a_{1,1}\mathbf{D}_y\mathbf{P}_t + a_{1,2}\mathbf{D}_x\mathbf{P}_t + \boldsymbol{\varepsilon}_{u,t}$$

$$\mathbf{V}_t = b_{1,1}\mathbf{D}_x\mathbf{P}_t + b_{1,2}\mathbf{D}_y\mathbf{P}_t + \boldsymbol{\varepsilon}_{v,t} \quad (1.8)$$

Where $\mathbf{U}_t$ and $\mathbf{V}_t$ are the discrete estimate of the vectorized zonal and meridional velocity components at time t, **D** and **P** are discrete operators for the spatial derivatives, and $\boldsymbol{\varepsilon}$ indicates a random error vector expressing the process model uncertainty.

The stochastic form of the sea-level pressure anomaly P decomposition is:

$$P_t(x,y) = \mu + \Sigma_{k=1..N}\, \alpha_{k,t}\, \phi_k(x,y) \quad (1.9)$$

where μ is the mean sea level pressure and $\phi_k(x, y)$, k=1…N are spatial structure functions.

The last layer in the hierarchy of the model prescribes the probability distributions of the random coefficients or associated hyperparameters entering the data and process model equations, including $a_{1,1}$, $a_{1,2}$, $b_{1,1}$, $b_{1,2}$, $\alpha_{k,t}$. The distribution specification for the error terms $\boldsymbol{\varepsilon}$ comprises a spatially structured part consisting of wavelet bases weighted by random coefficient (whose formulation is similar to the spatial decomposition followed in the BHM developed within PREFACE WP10), and a Gaussian noise part. This exemplifies how knowledge about the process under study, including the properties of its error, can be implemented in the BHM framework in a transparent way.

Wikle et al. (2013) present a recent review on hierarchical statistical approaches in the field of oceanography, particularly focused on the use of BHM for identification, quantification and management of uncertainty in data and models of ocean processes. The review spans different applications of BHMs, including data assimilation and inverse modeling, and long-term forecasting.

BHMs have been receiving increasing attention also by the paleoclimate community interested in reconstructing climate state and variability before the instrumental period. The paleoclimatic reconstruction problem consists of inferring a target, latent climate process in both space and time conditional on the (incomplete) observed instrumental and proxy time series and other available covariates. Instrumental and proxy data typically have different uncertainties, relate differently with the target climate process, and display different spatial and temporal dependencies. The paleoclimatic reconstruction problem is therefore naturally suited for the Bayesian approach as a unifying, hierarchical space-time modeling framework.

Tingley et al. (2012) review reconstruction methods focused on Bayesian inference that are currently favored by paleo-climatologists, focusing on hierarchical statistical models and describing how the different levels should be specified to the purpose of modeling a space-time process and inferring it from a number of different noisy and incomplete data sources. The review provides key specifications of the proposed hierarchical statistical space-time models, including the space-time structure of the target climate process and the relationships between the statistical processes characterizing the data sources and the target process. The hierarchical structure (with data, process

and parameter levels, see (a-c)) of the model allows to clarify that the relationship between the data and the target process is distinct from the covariance structure of the process, and further distinguishes between model assumptions (often quite similar in different studies) and the employed inference techniques (typically different in different studies), therefore providing a cohesive framework for propagating uncertainty through an analysis. The reader is referred to Tingley et al. (2012) for a more detailed overview of possible paleoclimate applications of hierarchical statistical modelling.

2 – Report on the Bayesian hierarchical model

This section provides an update on the status of the work progress concerning the development of the Bayesian hierarchical modelling strategy for describing space-time error dependencies. A first model was built upon the dynamic linear model (DLM) concept: it currently works at the grid-point level and has been tested and applied to a single-model ensemble of decadal climate predictions. It allows for a structural decomposition of the drift into regional and local components, each further separated into long-term and seasonal components. The conditional estimation of the drift components and their evolution achieved within the Bayesian framework allows accounting for uncertainty in both observations (i.e., empirical estimates of the drift) and model Gaussian error parameters through sampling of the posterior distribution of associated variances by MCMC. A second model was built upon the concept of spatial analysis. This model works on a multi-model framework and focuses on the spatial features of the bias (in contrast to the DLM, which focuses on the temporal development of the drift): it quantifies an overall common bias that is obtained by synthesizing bias across the different climate models in a multi-model ensemble, and model-specific individual bias components that are characterized as non-stationary spatial fields, further determining each model's contribution to the overall bias.

The following sections, based on two manuscripts currently under review in Nature Communications (Zanchettin et al., 2016) and in Stochastic Environmental Research and Risk Assessment (Arisido et al., 2016), provide a description of the current DLM implementation (sections 2.1 and 2.2), present some results from the current application of the model (section 2.3), and provide a description of the Bayesian hierarchical model for spatial analysis of the bias (section 2.4). An outline of future work is provided in section 2.5.

2.1 *Statistical model for drift estimation*

At the process level, the drift $\Delta_i(t)$ at grid point *i* and integration month *t* is described as a combination of a local (grid-point) component $\Lambda_i(t)$, further splitted in non-seasonal and seasonal contributions ($\lambda_i(t)$ and $\lambda^s_i(t)$, respectively), and a large-scale or regional component $P_i(t)$, also further splitted in non-seasonal and seasonal contributions ($\rho_i(t)$ and $\rho^s_i(t)$, respectively):

$$\Delta_i(t) = \Lambda_i(t) + P_i(t) = \lambda_i(t) + \lambda^s_i(t) + \rho_i(t) + \rho^s_i(t) \quad (2.1)$$

The non-seasonal drift components are modelled as follows: $\lambda_i(t) = \lambda_i(t\text{-}1) + \varepsilon_i(t)$, i.e., as a first-order random walk with Gaussian error $\varepsilon_i(t) \sim N(0,W_\varepsilon)$; $\rho_i(t) = 2\rho_i(t\text{-}1) - \rho_i(t\text{-}2) + \delta_i(t)$, i.e., as a second-order random walk with Gaussian error $\delta_i(t) \sim N(0,W_\delta)$. The latter is equivalent to a cubic spline smoothing and implies longer-term dependencies, i.e., longer memory, for regional errors compared to local errors. The seasonal components are modelled using harmonic functions (Laine et al., 2014).

At the data level, $\Lambda_i(t)$ and $P_i(t)$ are observed through differences between hindcast and assimilated values (D) according to the following assumption:

$$D_i(t) = \Delta_i(t) + \gamma_i(t) = \Lambda_i(t) + P_i(t) + \gamma_i(t) \quad (2.2)$$

$$D^*_i(t) = P_i(t) + \zeta_i(t) \quad (2.3)$$

where $\gamma_i(t) \sim N(0,V)$, $\zeta_i(t) \sim N(0,V)$, and $D^*_i(t)$ is the area-weighted mean of the differences between hindcast and assimilated values in the neighbourhood of grid-point *i*. An 11x11 grid-point area was considered in the current application.

2.2 *Dynamic Linear Model implementation*

The formulation of the statistical model above allows for a straightforward implementation within the dynamic linear model (DLM) framework. DLMs are based on a state-space approach, i.e., unobservable state variables are used that allow direct modeling of the process (***Y***) generating the observed data (***Z***)(Brogan, 1974; Laine et al., 2014). DLMs have the general form:

$$\boldsymbol{Z}(t) = \mathbf{F}\, \boldsymbol{Y}(t) + \mathbf{v}(t) \quad (2.4)$$

$$\boldsymbol{Y}(t) = \mathbf{G}\, \boldsymbol{Y}(t\text{-}1) + \mathbf{w}(t) \quad (2.5)$$

where *t* is the discrete time variable representing, in our case, monthly values, $\boldsymbol{Z}(t)$ is a vector of *p* observations at time *t*, $\boldsymbol{Y}(t)$ is the underlying state vector of dimension *m*, **G** is the *m*x*m* system matrix, and **F** is the *m*x*p* observation matrix. We suppose that $\mathbf{v}(t) \sim N(0,\mathbf{V})$ and $\mathbf{w}(t) \sim N(0,\mathbf{W})$ are the observation and model Gaussian errors, respectively, that are serially and mutually uncorrelated. In this formulation, the covariance matrices **V** and **W** contain the model parameters $\boldsymbol{\Theta}$. If we suppose that the unknown parameters are random the DLM formulation can be seen as a BHM where (2.4) and (2.5) are typically referred to as observation equation and system equation, respectively.

We have applied the DLM to each grid-point *i* of the considered domain. Following (2.2) and (2.3), and having *p* values of $D_i(t)$ (i.e., having *p* hindcasts) the observation vector in each grid-point is defined as $\boldsymbol{Z}(t) = \{D_i^1(t), ..., D_i^p(t), D_i^{*1}(t), ..., D_i^{*p}(t)\}'$.

Following (2) and accounting just for the 12-month seasonality for both $\lambda(t)$ and $\rho(t)$, the state vector is defined as $\boldsymbol{Y}(t) = \{\lambda(t), \lambda^s(t), \lambda^{s*}(t), \rho(t), \rho(t\text{-}1), \rho^s(t), \rho^{s*}(t)\}'$. Following Laine et al. (2014), $\lambda^s(t)$ and $\rho^s(t)$ are harmonics with general form $a\cos(\pi/6t) + b\sin(\pi/6t)$, while $\lambda^{s*}(t)$ and $\rho^{s*}(t)$ are their respective conjugates with general form $-a\sin(\pi/6t) + b\cos(\pi/6t)$, and *a* and *b* are constants. Therefore, the dimension of the state vector is *m*=7.

The observation matrix **F** is defined following (2.2) and (2.3). In order to reduce computational requirements, we apply the Bayesian inference on one common parameter V for all observations and two parameters **W**: W_λ for $\lambda(t)$, $\lambda^s(t)$ and $\lambda^{s*}(t)$, and W_ρ for $\rho(t)$, $\rho^s(t)$ and $\rho^{s*}(t)$. We define weakly informative lognormal priors with parameters (0,1) for V, W_λ and W_ρ.

The sequential definition of the process model (having a conditional dependency only on the previous time step) allows to use the Kalman filter formulas for calculating the posterior distribution

(1.1) (Laine et al., 2014). We use a slice-sampler algorithm (Radford, 2003) to iteratively sample from the full posterior distribution of the unknown parameters $\boldsymbol{\Theta}$. Then, Kalman filter and Kalman smoother are used to iteratively sample the $\boldsymbol{Y}$ states along the so-obtained Monte Carlo Markov Chain to derive marginal distributions for each of the state components.

2.3 Structural decomposition of the drift

We report here an exemplary result obtained by the application of the DLM described in sections 2.1 and 2.2 to output from the MiKlip prototype system for decadal climate predictions, which is based on the low resolution version of the Max Planck Institute - Earth System Model (Marotzke et al., 2016). We use the MiKlip experiments based on full-field assimilation of the ORAS4 ocean reanalysis data (Balmaseda et al., 2013). The "r1" ensemble initialized at the end of each year between 1960 and 2000 was used in the main analysis. We refrain from showing more results as this work is currently under review in the scientific journal Nature Communications. The shown results aim at exemplifying how the Bayesian framework upon which the DLM is built allows to focus on different aspects of uncertainty in climate model errors. Figure 2.1 shows, for an exemplary grid-point in the Angola-Benguela front region, that the largest uncertainty to account for in the estimation of the drift stems from the data model parameter V (included in the observation equation (2.4)), while the smallest uncertainty stems from the W_ρ parameter. We tested the impact of using different input data to the dynamic linear model by constructing different hindcast ensembles, including sub-ensembles of the r1 ensemble used in the main analysis, as well as super-ensembles including multiple realizations for each hindcast. Input data affect only slightly W_λ, as seen by the largely overlapping distributions in Figure 2.1b. A more constrained estimation of V is generally achieved by increasing the size of the ensemble, with only negligible changes in the estimated mean. In contrast, the use of strongly reduced ensembles can produce major discrepancies in the parameter estimation (Figure 2.1a). For very large ensembles including multiple realizations of the same hindcast, the more constrained estimation of V contrasts the more uncertain estimation of W_ρ, which therefore also becomes relatively more important (Figure 2.1c). These results reveal the non-trivial dependency of different of different uncertainty sources for drift estimation on the quality and quantity of available empirical information about the drift.

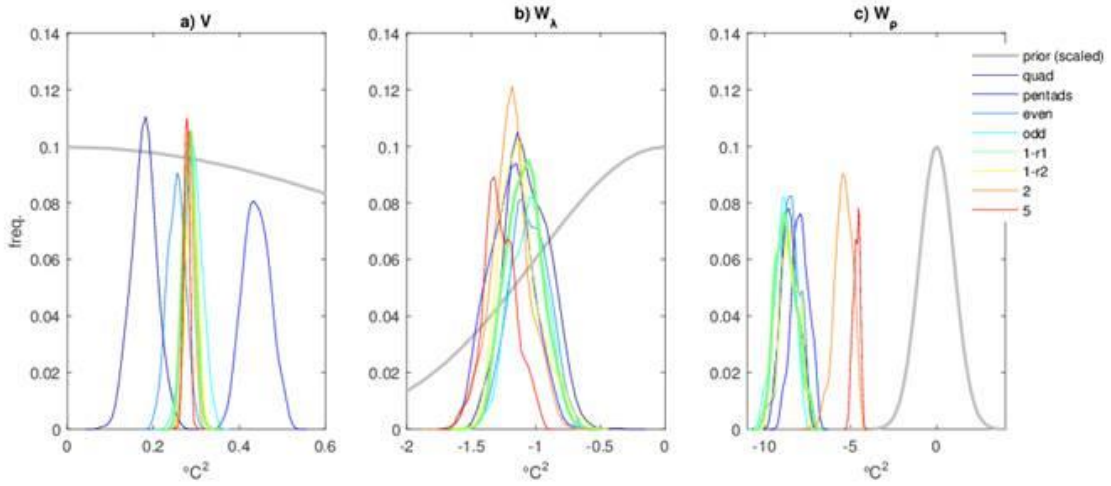

Figure 2.1 – Posteriori marginal log-distributions of the parameters of the Bayesian hierarchical model as sampled along the Monte Carlo Markov Chain. Shown parameters are the covariances of error components included in the dynamic linear model, for a) observational error, b) dynamical error for the local component, c) dynamical error for the regional component. Plotted estimates are for one exemplary grid-point over the Angola-Benguela front region, and different ensembles generated based on the available MiKlip ORAS4 hindcasts. The grey thick line is the scaled prior. The bold histograms indicate estimates for the setup used in the main analysis. Empirical distributions are smoothed with a 5-point Hanning window. uncertainty sources for drift estimation on the quality and quantity of available empirical information about the drift.

2.4 Bayesian hierarchical spatial assessment of the bias in a multi-model ensemble

The model is described in Arisido et al. (2016) and we report here a few technical notes on the general approach. Our aim is to obtain a statistical representation of climate model biases in a multi-model ensemble in order to separate an overall common bias from the individual components.

Climate model bias is determined by comparing output data against observations. We let Y(s) to represent the temperature observations and $X_j(s)$ to denote the temperature simulated by the climate model j in an ensemble of Q models, at the spatial location $s \in D$ for the domain D in R^2. Empirical climate model biases are then calculated as $B_j(s) = Y(s) - X_j(s)$; j = 1, ..., Q, where $B_j(s)$ denotes the bias of climate model j relative to the observations at spatial location s. For n sites in D, we observe the biases, namely $\{B_j(s_i), ..., B_j(s_n)\}$.

The Bayesian hierarchical model is formulated based on three levels: data, process, and parameters (Berliner et al., 2003, see also (1.1)). The data model captures the information given in the form of empirically measured biases, conditional on a hidden spatial bias process. The process level models the spatial structure and links the hidden spatial process to a set of parameters. In the parameter model, prior distributions are specified for the parameters. The three levels are specified in terms of probability distributions in a hierarchical structure shown in (a-b) of section 1 in this report.

2.4.1 Data model

We assume that the empirical bias $B_j(s)$ can be decomposed into two components: a spatial component $M_j(s)$ and a noise component $\varepsilon_j(s)$:

$$B_j(s) = M_j(s) + \varepsilon_j(s);\ j = 1,...,Q \qquad (2.6)$$

where $\{\varepsilon_j(s)\}$ is a Gaussian white noise with zero mean and variance $\sigma^2_{\varepsilon,j}$, and independent from the spatial component $\{M_j(s)\}$. Additionally, the noise component $\{\varepsilon_j(s)\}$ is assumed to be independent from $\{\varepsilon_k(s)\}$, for $k \neq j$. Thus, conditionally on the hidden spatial process $\{M_j(s)\}$, the observed bias $B_j(s)$ has a Gaussian distribution with mean $M_j(s)$, and variance $\sigma^2_{\varepsilon,j}$ that represents the data model level.

2.4.2 Process model

The spatial process $\{M(s)\}$, with $M(s) = (M_1(s), ..., M_Q(s))'$ is multivariate and can be modeled in different ways (Gelfand et al. 2010). We adopt an approach based on kernel basis functions (see, e.g., Higdon, 1998) and we suppose that:

$$M_j(s) = \sum_{k=1}^{p} \beta_{j,k} w_k(s) = \mathbf{w}(s)'\beta_j, \qquad (2.7)$$

with j=1, ... , Q, and where $\mathbf{w}(s) = \{w_1(s), ..., w_p(s)\}'$ is a vector of weighting kernels, $\beta_j = (\beta_{j,1}, ..., \beta_{j,p})'$ is a vector of unknown random parameters and $p << n$ denotes the number of components. We assume that the climate bias is an additive decomposition of a large scale error signal and small scale error signals including local model bias as well as local effects of, e.g., sampling of internal climate variability. The goal is to synthesize this overall common bias component, which is the same across all models in the ensemble. To this purpose we consider a random effect model (e.g., Furrer et al. 2007; Kang et al. 2012) for the random parameters β_j. More precisely, we assume that the k^{th} random parameter for the climate model j, $\beta_{j,k}$, is centered at the overall random effect α_k, namely:

$$\beta_{j,k} = \alpha_k + \nu_{j,k},\ \ j=1, ..., Q;\ k=1, ..., p \qquad (2.8)$$

The vector of the overall random effects $\alpha = (\alpha_1, ..., \alpha_p)'$ has multivariate Gaussian distribution $\alpha \sim Gau(\mathbf{0},\mathbf{G})$, where $\mathbf{G}$ is the non-diagonal pxp covariance matrix. The term $\nu_j = \{\nu_{j,1}, ..., \nu_{j,p}\}'$ denotes a vector of independently distributed zero-mean Gaussian processes, $\nu_j \sim Gau(\mathbf{0}; \tau^2_j \mathbf{I}_p)$, where $\mathbf{I}_p$ is the pxp identity matrix. Centering $\beta_{j,k}$ about the overall random effects ak corresponds to our assumption that the various models share a common bias signal. Nevertheless, we expect departures of each climate model bias from the overall common bias, and this difference is reflected by the variance parameter τ^2_j. Specifically, different values of τ^2_j across the various models indicate different levels of departure from the common bias. Alternatively, similar values of τ^2_j for different models indicate that they vary similarly about the overall common bias, suggesting the contribution of each climate model in estimating the overall common bias is similar. In fact, if we impose the restriction $\tau^2_1 = \tau^2_2 = ... = \tau^2_Q$, the common overall bias corresponds to a simple average of biases from all models. Combining (2.7) and (2.8) the process model is expressed as:

$$M_j(s) = \sum_{k=1}^{p} w_k(s)\left[\alpha_k + v_{j,k}\right] = \mu(s) + \eta_j(s) \qquad (2.9)$$

where $\mu(s) = \Sigma^{p}_{k=1,\ldots,p}\, w_k(s)\alpha_k$ specifies the overall common bias and $\eta_j(s) = w_k^*(s)\, v_{j,k}$ describes the jth model-specific features, or the departure of the jth model bias from the common overall bias. Here, we make a distinction between the weighting kernels used to describe $\mu(s)$ and η_j, i.e., w is different from w*. Since the individual components $\eta_j(s)$ aim to capture local-scale features, a larger number of kernels is required to capture this bias component compared to that necessary to describe the overall common bias, i.e., p < p'. Further, $\eta_j(s)$ follows the zero mean Gaussian distribution, $\eta_j(s) \sim$ Gau(**0**, τ^2_j**w***(s)**w***(s)'). In other words, the model suggests that the spatial process $M_j(s)$ is decomposed into an overall common component $\mu(s)$ and an individual component $\eta_j(s)$.

As an exemplary illustration of an application of the model, Figure 2.2 illustrates the posterior samples of the variance parameters $\{\tau^2_j: j=1,\ldots,6\}$ obtained for an application of the BHM to the case of near-surface air temperature bias over the tropical Atlantic and surrounding regions in an ensemble of six historical climate simulations performed with different models.

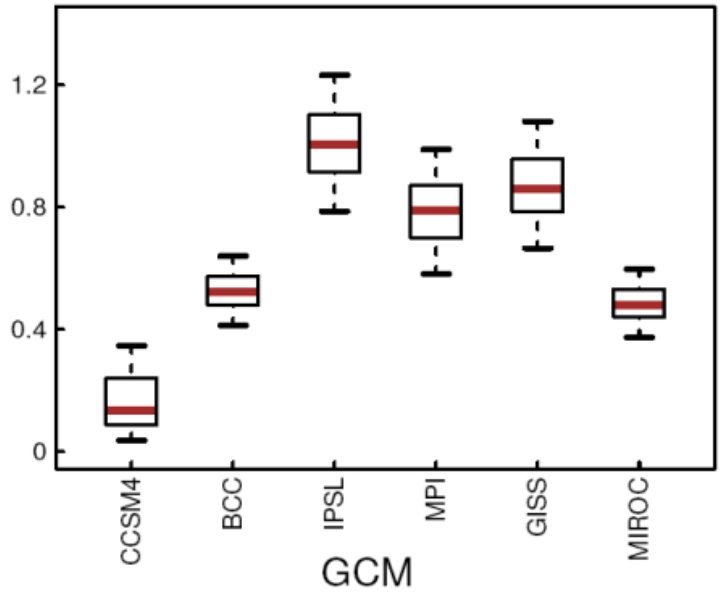

Figure 2.2 – Boxplots of the posterior samples of the variance parameters $\{\tau^2_j: j=1,\ldots,6\}$ (here illustrated as standard deviations) obtained for an application of the Bayesian model described in section 2.4 to the case of near-surface air temperature bias over the tropical Atlantic and surrounding regions. The bias data were calculated from 1950-1999 climatologies of six historical simulations contributing to CMIP5 described in Zanchettin et al. (2015) and of NCAR reanalysis data. Bold solid horizontal lines are the medians, the boxes indicate the 25th-75th percentile range, and the wiskers are the 5th-95th percentile ranges. Figure adapted from Arisido et al. (2016)

The variance parameters τ^2_j are useful to assess how each climate model bias varies about the overall common bias. The marked differences in terms of τ values across the individual climate models highlight how differently they contribute to the overall common bias. For instance, CCSM4 varies the

least, whereas IPSL and GISS vary the most about the overall common bias. Thus, in terms of weighting the contributions of each model in synthesizing the overall common bias, CCSM4 is ranked first, whereas IPSL and GISS have smaller weights. One benefit of the Bayesian hierarchical method is, therefore, that it allows to determine the heterogeneity across the climate models, highlighting the limitations of equal weight assumption often adopted by traditional empirical estimates.

2.5 Outline of future work

The core of the DLM described in sections 2.1 and 2.2 will be used to build a Bayesian hierarchical model for the climate model drift in SSTs in the Angola-Benguela front region, to test hypotheses about remote influence on the error development. This version of the DLM will make use of spatially-aggregated climatic indices representing regional features of the drift in key geophysical parameters, including shortwave and latent-heat surface fluxes in the Benguela region, winds in the equator/Benguela region, SSTs in the equatorial Pacific and in the Southern Ocean. The scope is to account for predictors affecting the evolution of the drift, and therefore move from a diagnostic perspective to a prognostic framework. The use of spatially-aggregated data for all quantities in the model, will allow to increase the complexity of the DLM, for instance to account for dynamic **F** and **G** operators in the observation and process equations in (2.3 and 2.4).

The BHM for probabilistically assessment and quantification of spatially referenced climate model biases in multi-model ensembles described in section 2.4 will be extended to include a temporal component. The main challenge will be to formulate a computationally efficient method for such an extensive approach taking into account the spatial and temporal features simultaneously.

Further development of the work is described in PREFACE Milestone MS34.

3 – Description of statistical models for spatial misalignment

In the current implementation of the DLM described in sections 2.1 and 2.2, observed and simulated data are expected to be given on the same grid. This expectation is fulfilled in our applications of the model so far, as we considered the output of the assimilation run used to initialize the hindcasts as our analog of observations. Also, in the current implementation of the BHM described in section 2.4, all model output is spatially interpolated to the same observational grid to resolve the misalignment between observations and outputs before fitting the BHM. The uncertainty associated to the interpolation can nonetheless affect the bias estimation in case of strong spatial misalignment between observations and model outputs. In our applications so far, the high spatial resolution of both observations and climate model output supported our expectation that the interpolation only minimally influences the result. However, when there is concern of substantial uncertainty due to interpolation, it may be desirable to build a model that is able to handle such spatial misalignment directly. One possible approach in the Bayesian framework is the hierarchical approach for nested block-level realignment (e.g., Banerjee et al. 2014), but this method requires that the model output to be nested in the observational grid (Mugglin and Carlin 1998). A simpler solution for the spatial model described in section 2.4 is, once the model outputs are firstly predicted to the observational grid using a stochastic model based approach such as the kriging method, to rectify the uncertainty that has been introduced by inflating the variance of the error $\varepsilon_j(s)$ in the data level (see equation 2.6 in section 2.4). We denote the predicted value from climate model j at spatial location s by $X^*_j(s)$. Its variance, $\delta^2_j(s) = \mathrm{var}(X^*_j(s))$ is zero if the model output grid and the observational grid coincide in s, otherwise it will be positive. Thus we specify

$$\mathrm{var}(\varepsilon_j(s)) = \sigma^2_j + \gamma_j \delta^2_j(s) \qquad (3.1)$$

where the modulating parameter γ_j is positive. This slight modification add further parameters to the BHM, for which we can assign a prior distributions similarly to σ^2.

Kriging is therefore a key component of our proposed approach to solve and accounts for spatial observations-simulations misalignments (for details on the method see: Banerjee et al., 2014). We developed a code in R that uses kriging solves spatial misalignment between model output and observations using the R package and the R library geoR (Ribeiro Jr and Diggle, 2001). The proposed software performs predictions based on maximum likelihood estimates, but a Bayesian approach can be easily implemented as well (see below). Different covariance models are tested, and the best performing model according to the Akaike Information Criterion is used for the prediction. Appendix 1 provide the R code and test input data in text format. Kriging allows also specifications of uncertainty associated to predictive distributions for prediction locations (observational grid) through posterior distributions within a Bayesian approach. In this case, let X(s) be the output of a climate model and Y(s) the observation in a site $s \in D \subset R^2$. We assume that observations $\mathbf{Y} = \{Y(s_1),...,Y(s_n)\}$ and the climate output $\mathbf{X^*} = \{X^*(s^*_1),...,X^*(s^*_m)\}$ are on different grids. The interest is therefore to predict $\mathbf{X} = \{X(s_1),...,X(s_n)\}$ using the information carried by $\mathbf{X^*}$. In probability terms it means that the conditional distribution $pr(\mathbf{X}|\mathbf{X^*},\boldsymbol{\Theta})$ is required.

This predictive distribution takes into account the parameter uncertainty by averaging over the parameter space the conditional distribution $pr(\mathbf{X}|\mathbf{X}^*,\boldsymbol{\Theta})$, with weights given by the posterior distribution for the model parameters $pr(\boldsymbol{\Theta}|\mathbf{X}^*)$. To measure the uncertainty, we require the prediction variance that is given by $Var(\mathbf{X}|\mathbf{X}^*)$. The geoR routine developed for R provides a way to compute $pr(\mathbf{X}|\mathbf{X}^*)$ and $Var(\mathbf{X}|\mathbf{X}^*)$. Figure 3.1 illustrates an exemplary idealized application of the Bayesian kriging to the case of predicting SST in a number of sites where ecological observations are available (in this case fish eggs and larvae based on the ichthyoplankton dynamics; see Lett et al., 2008 for a more detailed description). The original SST data are on a regular grid (left panel), while the prediction sites are irregularly distributed (mid panel). The method yields an uncertainty estimation for each prediction location (right panel).

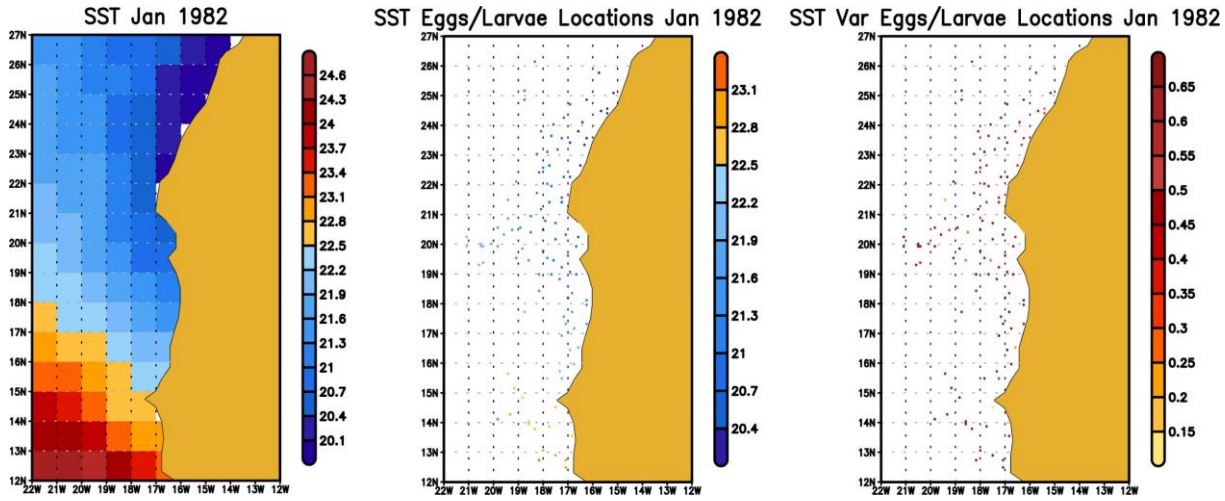

Figure 3.1 - Results from an exemplary idealized application of the kriging method to the case of SST misalignment: observed values are on a regular grid (left panel), but information is requested on an irregular grid, misaligned from the observational grid; kriging is used to predict the SST values on the irregular grid (middle panel) and quantify the uncertainty associated to the prediction (right panel).

References

Arisido, M. W., C. Gaetan, D. Zanchettin, A. Rubino (2016) Assessing climate model biases: A Bayesian hierarchical approach for the analysis of multi-model output. Under review in Stoch. Environm. Res. Risk. Ass. (ms: SERR-D-16-00234)

Balmaseda, M. A., Mogensen, K. and Weaver, A. T. (2013), Evaluation of the ECMWF ocean reanalysis system ORAS4. Q.J.R. Meteorol. Soc., 139: 1132–1161. doi: 10.1002/qj.2063

Banerjee, S., Carlin, B.P., Gelfand, A.E. (2014) Hierarchical modeling and analysis for spatial data. CRC Press, New York

Berliner, L. M. (1996). Hierarchical Bayesian time series models. In: Maximum Entropy and Bayesian Methods (Santa Fe, NM, 1995). Fund. Theories Phys., 79:15–22, Kluwer Academic, Dordrecht. MR1446713

Berliner, L.M., Kim, Y. (2008) Bayesian design and analysis for superensemble-based climate forecasting. J. Climate 21: 1891–1910.

Berliner, L.M., Milliff, Wikle, C.K. (2003) Bayesian hierarchical modeling of air-sea interaction. J. Geophys. Res. 108 (C4), 3104. http://dx.doi.org/10.1029/2002JC001413

Brogan, W. L. (1974). Modern Control Theory (3rd International edition). Prentice-Hall, ISBN-10:0135904153, 736 pp.

Buser, C. M., Künsch, H. R., Lüthi, D., Wild, M., and Schär, C. (2009). Bayesian multi-model projection of climate: bias assumptions and interannual variability. Clim. Dyn., 33:849-868

Christensen, W.F., Sain S.R. (2012) Latent variable modeling for integrating output from multiple climate models. Mathematical geosciences 44:395–410.

Duan, Q.m Phillips, T. J. (2010) Bayesian estimation of local signal and noise in multimodel simulations of climate change. J. Geophys. Res. Atmos. 115(D18), doi:10.1029/2009JD013654

Furrer, R., S. R. Sain, D. Nychka, C. Tebaldi, and G. A. Meehl (2007) Multivariate Bayesian Analysis of Atmosphere–Ocean General Circulation Models. Environm. Ecol. Stat., 14:249-266

Giorgi, F., Mearns, L.O. (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from aogcm simulations via the reliability ensemble averaging (REA) method. Journal of Climate 15(10):1141–1158.

Higdon, D. (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean. Environmental and Ecological Statistics 5:173-190

Ho, C. K., Stephenson, D. B., Collins, M., Ferro, C. A. T., and Brown, S. J. (2012). Calibration Strategies: A Source of Additional Uncertainty in Climate Change Projections. Bulletin of the American Meteorological Society, 93:21-26

Jun, M., R. Knutti, and D. W. Nychka (2008) Spatial Analysis to Quantify Numerical Model Bias and Dependence: How Many Climate Models Are There? J. Amer. Stat. Ass., 103, 483, doi:10.1198/016214507000001265

Kang, E. L., Cressie, N., and Sain, S. R. (2012). Combining outputs from the North American Regional Climate Change Assessment Program by using a Bayesian hierarchical model. Journal of the Royal Statistical Society: Series C (Applied Statistics), 61:291-313

Knutti, R. (2010). The end of model democracy? Climatic change 102(3-4):395–404.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., Meehl, G.A. (2010) Challenges in combining projections from multiple climate models. Journal of Climate 23(10):2739–2758.

Laine, M., N. Latva-Pukkila, and E. Kyrölä (2014) Analysing time-varying trends in stratospheric ozone time series using the state space approach. Atmos. Chem. Phys., 14, 9707–9725, doi:10.5194/acp-14-9707-2014

Lett, C., Verley, P., Mullon, C., Parada, C., Brochier, T., Penven, P., Blanke, B. (2008) A Lagrangian tool for modelling ichthyoplankton dynamics. Environmental Modelling & Software 23, 1210–1214.

Li, B., D. W. Nychka, C. M. Ammann (2010) The Value of Multiproxy Reconstruction of Past Climate. J. Am. Stat. Ass. Maraun, D. (2012). Nonstationarities of regional climate model biases in European seasonal mean temperature and precipitation sums. Geophysical Research Letters, 39:L06706

Marotzke, J., W. A. Müller, F. S. E. Vamborg, P. Becker, U. Cubasch, H. Feldmann, F. Kaspar, C. Kottmeier, C. Marini, I. Polkova, K. Prömmel, H. W. Rust, D. Stammer, U. Ulbrich, C. Kadow, A. Köhl, J. Kröger, T. Kruschke, J. G. Pinto, H. Pohlmann, M. Reyers, M Schröder, F. Sienz, C. Timmreck, M. Ziese, 2016: MiKlip - a National Research Project on Decadal Climate Prediction. Bull. Amer. Meteor. Soc. doi:10.1175/BAMS-D-15-00184.1

Milliff, R. F., A. Bonazzi, C. K. Wikle, N. Pinardi, and L. M. Berliner (2011) Ocean ensemble forecasting. Part I: Ensemble Mediterranean winds from a Bayesian hierarchical model. Q. J. R. Meteorol. Soc. 137: 858–878

Mugglin, A.S., Carlin, B.P. (1998) Hierarchical modeling in Geographic Information Systems: population interpolation over incompatible zones. Journal of Agricultural, Biological and Environmental Statistics 3: 111130.

Neeley, E. S., W. F. Christensen and S. R. Sain (2014) A Bayesian spatial factor analysis approach for combining climate model ensembles. Environm., doi: 10.1002/env.2277

Radford, N. M. (2003) Slice Sampling. Ann. Stat. 31(3): 705–767

Reilly, J., P. H. Stone, C. E. Forest, M. D. Webster, H. B. Jacoby, and R. G. Prinn (2001), Uncertainty in climate change assessments, Science, 293, 430–433, doi:10.1126/science.1062001

Ribeiro Jr., P.J., and Diggle, P.J. (2001) geoR: A package for geostatistical analysis. R-NEWS Vol 1, No 2. ISSN 1609-3631. Available for download at: http://cran.r-project.org/doc/Rnews

Robert, C.P. and Casella, G (2004) Monte Carlo Statistical Methods, Springer, New York

Royle, J.A., Berliner, L.M., Wikle, C.K., Milliff, R.F. (1998) A hierarchical spatial model for constructing surface winds from scatterometer data in the Labrador Sea, in: Case Studies in Bayesian Statistics IV, C. Gatsonis, R. E. Kass, A. Cariquiry, B. Carlin, A. Cariquiry, A. Gelman, I. Verdinelli, and M. West (Eds.), Springer-Verlag: 367–381.

Sain, S.R., Furrer R. (2009) Combining climate model output via model correlations. Stochastic Environmental Research and Risk Assessment 24:821–829.

Salazar, E., Sansó, B., Finley, A.O., Hammerling, D., Steinsland, I., Wang, X., Delamater, P. (2011) Comparing and blending regional climate model predictions for the American southwest. Journal of Agricultural, Biological, and Environmental Statistics 16(4):586–605.

Sang, H., Jun M., Huang J.Z. (2011) Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors. The Annals of Applied Statistics 5(4):2519–2548.

Smith, R. L., Tebaldi, C., Nychka, D., and Mearns, L. (2009), "Bayesian Modeling of Uncertainty in Ensembles of Climate Models," Journal of the American Statistical Association, 104: 485, doi:10.1198/jasa.2009.0007

Tebaldi C, Sanso B. (2009) Joint projections of temperature and precipitation change from multiple climate models: a hierarchical Bayesian approach. Journal of the Royal Statistical Society, Series A 172:83–106

Tebaldi, C., and Knutti, R. (2007), "The Use of the Multi-Model Ensemble in Probabilistic Climate Projections," Philosophical Transactions of the Royal Society, Ser. A, 365, 2053–2075

Tebaldi, C., Smith, R. L., Nychka, D., and Mearns, L. O. (2005). Quantifying Uncertainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multimodel Ensembles. Journal of Climate, 18:1524-1540

Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt, B. Rajaratnam (2012) Piecing together the past: statistical insights into paleoclimatic reconstructions. Quat. Sc. Rev. 35:1-12

Wikle, C. K., R. F. Milliff, R. Herbei, W. B. Leeds (2013) Modern Statistical Methods in Oceanography: A Hierarchical Perspective. Statistical Science, 28:4, 466–486, doi:10.1214/13-STS436

Wikle, C.K., Milliff, R.F., Nychka, D., Berliner, L.M. (2001) Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. J. Amer. Statist. Assoc. 96: 382–397.

Zanchettin, D., Bothe, O., Lehner, F., Ortega, P., Raible, C. C., and Swingedouw, D. (2015) Reconciling reconstructed and simulated features of the winter Pacific/North American pattern in the early 19th century, Clim. Past, 11, 939-958, doi:10.5194/cp-11-939-2015

Zanchettin D., Arisido M. W., Gaetan C., Modali K., Rubino A. (2016) Structural decomposition of climate model drifts: A Bayesian approach. Under review in Nat. Comm. (ms: NCOMMS-16-22293)

Appendix 1

R-code to solve spatial misalignment between observations and model output:

```
#=============================================================================
# Software for solving spatial misalignment between model output and observations
# using the R package and the library geoR.
# The misalignment is solved using kriging. The software performs predictions based on
# maximum likelihood estimates, but a Bayesian approach can be easily implemented as
# well. Different covariance models are tested, and the best performing model according
# to the Akaike Information Criterion is used for the prediction.
# The code requires:
# R and geoR installed
# input model output (longitude, latitude, data)
# a list of spatial locations (longitude, latitude) where to perform the prediction
#
# To install geoR type in the console (then follow instructions):
# install.packages("geoR")
#
# Original code written by: Carlo Gaetan; modified by: Davide Zanchettin
# Last update: 19/19/2016
# ============================================================================

rm(list = ls())
library(geoR)
# == 1 ========= LOAD AND PRE-PROCESS DATA ================
# In this test case we use ASCII files for all inputs, this can be easily changed to
# handle otherfile formats, such as .nc
# load model output and associated spatial information:
data.model<-read.table(file="test_data_model.txt",header = TRUE)
# load the prediction coordinates
pred.grid<-read.table(file="test_pred_coords.txt",header = TRUE)
# Convert the model data in a geodata object
data.geo<-as.geodata(data.model)
# supervised assessment of the data spatial structure is necessary to appropriately
# define the kriging model:
# we plot the data to check for possible non stationarity
plot(data.geo)
# non-stationarity is confirmed inspecting the empirical variogram
# in the four main directions
plot(variog4(data.geo))
# we accordingly define a trend to be removed
trend <-"1st" # other choices: constant "cte", 2nd order "2nd"
# == 2 ========= DEFINE MODEL PARAMETERS FOR THE KRIGING ================
# We search for a covariance model among the most frequently used models
# User can change/expand the list, to check for available models type
# help(cov.spatial) in the console
cov.model<-c("exponential","spherical","cubic")
nmod<-length(cov.model) # the number of models that we will compare
# We estimate covariance parameters by fitting a parametric model using the
# maximum likelihood method.
# First of all we calculate the initial guess for the covariance parameters:
# the variance ....
fit.lm<-lm(data.geo$data~data.geo$coords)
sigma2.hat<-mean(residuals(fit.lm)^2)
# ... and the range
```

```
phi.data<-quantile(dist(data.geo$coords),0.5)
# We create a matrix for storing the estimates and statistics from the fitting
results<-matrix(0,nmod,3+length(coef(fit.lm)))
# fitting procedure, testing all the nmod models available
for (i in 1:nmod) {
 fit<-likfit(data.geo,trend = trend,ini.cov.pars = c(sigma2.hat,phi.data),cov.model=cov.model[i])
 results[i,1]<-fit$AIC
 results[i,2:3]<-fit$cov.pars
 results[i,-(1:3)]<-fit$beta
}
# determine the best model according to the AIC:
best.model<-which.min(results[,1])
# eventually define the parameters for the kriging
krige.param<-krige.control(type.krige = "ok",trend.d=trend,trend.l = trend,
                           beta=results[best.model,-(1:3)],
                           cov.model = cov.model[best.model],
                           cov.pars=results[best.model,(2:3)])
# = 3 ========= PERFORM THE KRIGING AND SAVE PREDICTIONS ================
# predict values at the wanted locations
pr <- krige.conv(data.geo, loc=pred.grid, krige=krige.param)
# save the results with an indication of the best model
write.table(cbind(pred.grid,pr$predict,pr$krige.var),
            file=paste("predictions-from-", cov.model[best.model],"-model.txt",sep=""),
            row.names = FALSE,col.names = c("lon","lat","prediction","variance"))

# ====================================================================
# ====================================================================
```

Test input files to run the R code (copy paste to a text editor to create the files)

test_pred_coords.txt

```
"lon.o" "lat.o"
-18 13.061197
-16.833333 13.548019
-18.833333 14.841303
-17 15.645769
-18.75 17.722389
-16.833333 17.722389
-20.416667 19.382525
-18.416667 18.277542
-16.666667 19.775386
-21.083333 20.24555
-18.416667 20.636276
-17.583333 20.636276
-18.25 23.035973
-16.833333 22.651663
-18.75 25.167614
-17.583333 24.030053
-15.75 24.18232
-18.333333 26.145044
```

test_data_model.txt

```
"lon.m" "lat.m" "data.m"
-24.5 34.5 18.145401
-23.5 34.5 18.078917
-22.5 34.5 18.027298
-21.5 34.5 17.987389
-20.5 34.5 17.940502
-19.5 34.5 17.886814
-18.5 34.5 17.896622
-17.5 34.5 17.959257
-16.5 34.5 17.95348
-15.5 34.5 17.90065
-14.5 34.5 17.91044
-13.5 34.5 17.990269
-12.5 34.5 17.982979
-11.5 34.5 17.870398
-10.5 34.5 17.771492
-9.5 34.5 17.663475
-8.5 34.5 17.607363
-7.5 34.5 17.631123
-6.5 34.5 17.548195
-24.5 33.5 18.638432
-23.5 33.5 18.615002
-22.5 33.5 18.57058
-21.5 33.5 18.506702
-20.5 33.5 18.4361
-19.5 33.5 18.353292
-18.5 33.5 18.365152
-17.5 29.5 20.016197
-16.5 29.5 19.886494
-15.5 29.5 19.654039
-14.5 29.5 19.426239
-13.5 29.5 19.19722
-12.5 29.5 19.024338
-11.5 29.5 18.968414
-10.5 29.5 18.779444
-24.5 28.5 20.722494
-23.5 28.5 20.717613
-22.5 28.5 20.657736
-21.5 28.5 20.544022
-20.5 28.5 20.452877
-19.5 28.5 20.375641
-18.5 28.5 20.325451
-17.5 28.5 20.308908
-16.5 28.5 20.190445
-15.5 28.5 19.958775
-14.5 28.5 19.723335
-13.5 28.5 19.515896
-12.5 28.5 19.145676
-11.5 28.5 19.071997
-24.5 27.5 21.09367
-23.5 27.5 21.147472
-22.5 27.5 21.115993
-21.5 27.5 20.99926
-20.5 27.5 20.869976
-19.5 27.5 20.722078
-18.5 27.5 20.625551
-17.5 27.5 20.58046
-16.5 27.5 20.419743
-15.5 27.5 20.1493
-14.5 27.5 19.924217
-13.5 27.5 19.686445
-24.5 26.5 21.473385
-23.5 26.5 21.569937
-22.5 26.5 21.572628
-21.5 26.5 21.473003
-20.5 26.5 21.311937
-19.5 26.5 21.084301
-18.5 26.5 20.901493
-17.5 26.5 20.759575
-16.5 26.5 20.500969
-15.5 26.5 20.179893
-14.5 26.5 20.005894
-24.5 25.5 21.829956
-23.5 25.5 21.880066
-22.5 25.5 21.850639
-21.5 25.5 21.725945
-20.5 25.5 21.54113
-19.5 25.5 21.286901
-18.5 25.5 21.020136
-17.5 25.5 20.74387
-23.5 17.5 23.106407
-22.5 17.5 22.977242
-21.5 17.5 22.653553
-20.5 17.5 22.470512
-19.5 17.5 22.393433
-18.5 17.5 22.084312
-17.5 17.5 21.718027
-16.5 17.5 21.589159
-24.5 16.5 23.348642
-23.5 16.5 23.374643
-22.5 16.5 23.257502
-21.5 16.5 22.941814
-20.5 16.5 22.745018
-19.5 16.5 22.615131
-18.5 16.5 22.349901
-17.5 16.5 22.136673
-16.5 16.5 22.080067
-24.5 15.5 23.599804
-23.5 15.5 23.575127
-22.5 15.5 23.497173
-21.5 15.5 23.293831
-20.5 15.5 23.106049
-19.5 15.5 22.866341
-18.5 15.5 22.543568
-17.5 15.5 22.368332
-16.5 15.5 22.297155
-24.5 14.5 23.818783
-23.5 14.5 23.702259
-22.5 14.5 23.703911
-21.5 14.5 23.726139
-20.5 14.5 23.582172
-19.5 14.5 23.203403
-18.5 14.5 22.75758
-17.5 14.5 22.549826
-24.5 13.5 24.04789
-23.5 13.5 23.962259
-22.5 13.5 24.04623
-21.5 13.5 24.180634
-20.5 13.5 24.110043
-19.5 13.5 23.777782
-18.5 13.5 23.262775
-17.5 13.5 22.8829
-16.5 13.5 22.889908
-24.5 12.5 24.304413
-23.5 12.5 24.356556
-22.5 12.5 24.507935
-21.5 12.5 24.635487
-20.5 12.5 24.680607
-19.5 12.5 24.599907
-18.5 12.5 24.08736
-17.5 12.5 23.403912
-16.5 12.5 23.191395
-24.5 11.5 24.66238
```

-17.5 33.5 18.463032	-16.5 25.5 20.387247	-23.5 11.5 24.776762
-16.5 33.5 18.461884	-15.5 25.5 20.084164	-22.5 11.5 24.951927
-15.5 33.5 18.385132	-14.5 25.5 20.024914	-21.5 11.5 25.081587
-14.5 33.5 18.372902	-24.5 24.5 22.142166	-20.5 11.5 25.183598
-13.5 33.5 18.431051	-23.5 24.5 22.066067	-19.5 11.5 25.224955
-12.5 33.5 18.386391	-22.5 24.5 21.941641	-18.5 11.5 24.778709
-11.5 33.5 18.214281	-21.5 24.5 21.75272	-17.5 11.5 24.045332
-10.5 33.5 18.070671	-20.5 24.5 21.550884	-16.5 11.5 23.907455
-9.5 33.5 17.93914	-19.5 24.5 21.320679	-15.5 11.5 24.792473
-8.5 33.5 17.70793	-18.5 24.5 20.983046	-24.5 10.5 25.106297
-7.5 33.5 17.694485	-17.5 24.5 20.574417	-23.5 10.5 25.198812
-24.5 32.5 19.086906	-16.5 24.5 20.209457	-22.5 10.5 25.349983
-23.5 32.5 19.136896	-15.5 24.5 20.00679	-21.5 10.5 25.48774
-22.5 32.5 19.116413	-24.5 23.5 22.339176	-20.5 10.5 25.576653
-21.5 32.5 19.030277	-23.5 23.5 22.14846	-19.5 10.5 25.589922
-20.5 32.5 18.944124	-22.5 23.5 21.964172	-18.5 10.5 25.246351
-19.5 32.5 18.846058	-21.5 23.5 21.768909	-17.5 10.5 24.679815
-18.5 32.5 18.842823	-20.5 23.5 21.591087	-16.5 10.5 24.673515
-17.5 32.5 18.934465	-19.5 23.5 21.410828	-15.5 10.5 25.224014
-16.5 32.5 18.913292	-18.5 23.5 20.988657	-14.5 10.5 26.167328
-15.5 32.5 18.797447	-17.5 23.5 20.41778	
-14.5 32.5 18.723894	-16.5 23.5 20.101553	
-13.5 32.5 18.691895	-24.5 22.5 22.42445	
-12.5 32.5 18.575626	-23.5 22.5 22.130312	
-11.5 32.5 18.355335	-22.5 22.5 21.930611	
-10.5 32.5 18.212433	-21.5 22.5 21.79888	
-9.5 32.5 18.155609	-20.5 22.5 21.705446	
-24.5 31.5 19.517925	-19.5 22.5 21.629299	
-23.5 31.5 19.559135	-18.5 22.5 21.135151	
-22.5 31.5 19.532757	-17.5 22.5 20.458681	
-21.5 31.5 19.444714	-16.5 22.5 20.060604	
-20.5 31.5 19.37141	-24.5 21.5 22.511827	
-19.5 31.5 19.296373	-23.5 21.5 22.186901	
-18.5 31.5 19.280172	-22.5 21.5 22.001911	
-17.5 31.5 19.333891	-21.5 21.5 21.919596	
-16.5 31.5 19.280769	-20.5 21.5 21.828867	
-15.5 31.5 19.12685	-19.5 21.5 21.712351	
-14.5 31.5 18.998377	-18.5 21.5 21.187546	
-13.5 31.5 18.887627	-17.5 21.5 20.608162	
-12.5 31.5 18.730865	-24.5 20.5 22.638256	
-11.5 31.5 18.526865	-23.5 20.5 22.363331	
-10.5 31.5 18.376181	-22.5 20.5 22.221588	
-9.5 31.5 18.330311	-21.5 20.5 22.174057	
-24.5 30.5 19.965342	-20.5 20.5 22.003967	
-23.5 30.5 19.914927	-19.5 20.5 21.699444	
-22.5 30.5 19.851475	-18.5 20.5 21.180746	
-21.5 30.5 19.777428	-17.5 20.5 20.709686	
-20.5 30.5 19.742561	-16.5 20.5 20.584667	
-19.5 30.5 19.732014	-24.5 19.5 22.757185	
-18.5 30.5 19.709419	-23.5 19.5 22.576271	
-17.5 30.5 19.690987	-22.5 19.5 22.450441	
-16.5 30.5 19.582092	-21.5 19.5 22.345987	
-15.5 30.5 19.375612	-20.5 19.5 22.146904	
-14.5 30.5 19.187035	-19.5 19.5 21.840584	
-13.5 30.5 19.007244	-18.5 19.5 21.384506	
-12.5 30.5 18.856388	-17.5 19.5 20.937256	
-11.5 30.5 18.757885	-16.5 19.5 20.769146	
-10.5 30.5 18.635067	-24.5 18.5 22.886948	

-9.5 30.5 18.409502 -24.5 29.5 20.366877 -23.5 29.5 20.299608 -22.5 29.5 20.2244 -21.5 29.5 20.140738 -20.5 29.5 20.099159 -19.5 29.5 20.088842 -18.5 29.5 20.056791	-23.5 18.5 22.837582 -22.5 18.5 22.701916 -21.5 18.5 22.445381 -20.5 18.5 22.257753 -19.5 18.5 22.117559 -18.5 18.5 21.745264 -17.5 18.5 21.301907 -16.5 18.5 21.134481 -24.5 17.5 23.08506	