Development of spatio-temporal models of fishery catch-per-unit-effort data to derive indices of relative abundance

La Jolla, CA, USA, February 26-March 2, 2018.

Agenda

Monday

9:00 course on VAST

11:00 Welcome – Gerard DiNardo

11:15 Introduction to the issues – Rick Methot

12:00 Lunch

1:00 Basic concepts – Jim Thorson

2:00 Statistical issues - Hans Skaug

3:00 Coffee

3:30 Computational issues – Kasper Kristensen

4:30 Discussion on TMB based methods

5:00 End

Tuesday

9:00 am Non-fisheries perspective - Christopher Wikle

10:00 Coffee

10:30 Spatial autoregressive models - Jay Ver Hoef

11:30 Alternative methods - Nicole Augustin

12:30 Lunch

1:30 Discussion on alternative methods

2:30 Issues highlighted by applications - Toshihide Kitakado

3:30 Coffee

4:00 Likelihood functions for assessments – Mark Maunder

4:30 End

Wednesday

9:00 Francesca Forrestal

9:30 John Walter

10:00 Coffee

10:30 Best

11:00 Hoyle

11:30 Ducharme-Bath

12:00 Lunch

1:00 Siddeek

1:30 Discussion on CPUE issues

2:00 Xu

2:30 Kuriyama

3:00 coffee

3:30 Cao

4:00 Discussion on modeling composition data

4:30 End

6:00 Social event at SIO Surfside

Thursday

9:00 Conn

9:30 Torrejon

10:00 Coffee

10:30 Sculley

11:00 Deadman

11:30 Stock

12:00 Discussion on alternative methods

12:30 Lunch

1:30 Vert-pre

2:00 Monk

2:30 Perretti

3:00 Coffee

3:30 Tremblay-Boyer

4:00 Discussion on multi-species and covariates

4:30 General discussion/Application planning

5:00 End

Friday

9:00 – 5:00 Applications

Invited Abstracts

Rick Methot

Assessment implications for spatial analysis of cpue

A common practice in stock assessment modeling is to use the CPUE of a fishery as an index of population abundance. The age/length selectivity for this CPUE index is assumed to be the same as that of the fishery catch. The problem with this assumption is that the age composition of the fishery catch is catch-weighted, not abundance weighted. So if the fishery concentrates effort in a region that contains large fish, the catch age composition will contain a disproportionate amount of large fish and the estimated selectivity for the fishery will correctly show that the fishery selectivity is tilted towards large fish. However, if the CPUE is to correctly represent a quantity proportional to population abundance filtered through selectivity, then this CPUE selectivity needs to be responsive to region area x area-specific stock numbers, not regional catch. While it may be possible to process the raw length composition data once in a catch-weighted manner for the fishery catch and then in an abundance weighted manner for the CPUE index, this will result in double-use of the same data with corresponding correlation of observation errors in this data. Recommended solutions to this dilemma are sought from this workshop.

James Thorson

Spatio-temporal models for fishery catch and effort data: Solving old problems and creating new ones

Spatio-temporal models can resolve many big questions in fisheries science by estimating population abundance, local productivity, habitat associations, climate impacts, and ecosystem linkages. However, realizing the benefits of this approach worldwide requires methods for analyzing fishery catch and effort (CPUE) data because most marine ecosystems do not have data from planned surveys of marine communities.

In this talk, I first provide a brief overview of spatio-temporal concepts and parameter estimation, and introduce four questions that guide past and future development of spatio-temporal analysis for fishery CPUE. I then argue that a spatio-temporal approach addresses the two main questions regarding fishery CPUE analysis raised over the past decade of research:

1. How should we impute density (and predictive variance) in areas with little data? and
2. When can we use auxiliary data to separate changes in fishery catchability and fish density?

Specifically, a spatio-temporal approach can control for random changes in the spatial footprint of fishery data, and we can derive identifiability conditions for using survey, multispecies, and locational data to unconfound fishery catchability and fish density. I then introduce new questions that arise in a spatio-temporal approach:

3. How should we account for non-random selection of fishing locations? and

4. How should we process “biological data” (subsamples of age, length, or sex from fishery catch records) in conjunction with fishery CPUE?

I specifically summarize recent research regarding “preferential sampling” to argue that non-random selection of fishing locations is an important and persistent issue. I also present preliminary research suggesting that the standard approach to the spatial expansion of biological data will result in a mismatch between abundance index and compositional estimates generated from fishery CPUE data. Finally, I conclude by recommending that our community does a better job of advocating that available survey data be made public, digitizing historical logbook records when available, and publicizing trends in abundance for species worldwide.

Hans Skaug

Some new items in the space-time statistical toolbox that may be useful in fisheries

The R-packages "R-INLA" and "mgcv" allow spatial density surfaces to be fit to CPUE data. Their command line interfaces are flexible, but may still be too restrictive for the complicated data types that are commonly collected in fisheries. However, these R-packages contain elements/functionality that facilities the process of setting up spatial models. Models with "barriers", such as islands in the ocean, is one example. I will discuss how model objects from R-INLA and mgcv can be imported into TMB (Template Model Builder). Hence, TMB's flexibility and computational speed can be combined with the other packages’ easy-to-use interfaces for formulating the spatial components of the model. I will use CPUE data from North East Atlantic mackerel as an example.

Kasper Kristensen

Spatio-temporal models in TMB

Template Model Builder (TMB) is an open source R package that enables quick implementation of complex nonlinear random effects (latent variable) models in a manner similar to the AD Model Builder package (ADMB, Fournier et al. 2011). In addition, it offers easy access to parallel computations. The user defines the joint likelihood for the data and the random effects as a C++ template function, while all the other operations are done in R; e.g., reading in the data. The package evaluates and maximizes the Laplace approximation of the marginal likelihood where the random effects are automatically integrated out. This approximation, and its derivatives, are obtained using automatic differentiation (up to order three) of the joint likelihood.

This talk will focus on the capabilities of TMB to handle spatial and temporal latent variable models. It will be demonstrated how to implement such models and what output is generated.

Limitations of the approach when scaling to very large numbers of latent variables will be discussed. We will suggest possible solutions e.g. utilizing model properties beyond conditional independence structure to handle much larger models.

Christopher K. Wikle

Recent Advances in Quantifying Uncertainty in Spatio-Temporal Statistical Models

Spatio-temporal data are ubiquitous, and their study is important for understanding and predicting a wide variety of processes of interest to scientists, engineers, and policy makers. One of the primary difficulties in modeling spatial processes that change with time is the complexity of the dependence structures that must describe how such a process varies, and the presence of high-dimensional datasets and prediction domains. Much of the methodological development in recent years has considered either efficient moment-based approaches or spatio-temporal dynamical models. To date, most of the focus on statistical methods for dynamic spatio-temporal processes has been on linear models or highly parameterized nonlinear models. Even in these relatively simple models, there are significant challenges in specifying parameterizations that are simultaneously useful scientifically and efficient computationally. The hierarchical modeling framework can facilitate the incorporation of scientific information into these modeling structures, but at a computational cost. Approaches for nonlinear spatio-temporal data from outside statistics (e.g., analog methods, neural networks, agent-based models) offer intriguing alternatives in which the parameter space can be dramatically reduced. Yet, these methods often do not have formal mechanisms to quantify various sources of uncertainty in observations, model specification, and parameter estimation. This talk presents a brief overview of parametric statistical methods for spatio-temporal dynamical models as well as some recent attempts to place non-statistical models for such processes into a more rigorous uncertainty quantification framework. A common thread between all of these models is a “deep” modeling structure and the desire for efficient computation.

Nicole H. Augustin

Using generalised additive mixed models for estimating relative fish abundance in space and time

Generalized additive mixed models (GAMMs) provide a flexible tool to model fishery catch and effort data for estimating relative fish abundance in space and time. The estimation of GAMMs can be carried out using the R package mgcv. Often the objective of the modelling is to see whether there is evidence for a space-time interaction in abundance which may suggest stock-movement or local depletion. GAMMs allow for the inclusion of such a space-time interaction represented via a tensor product of a two dimensional smooth for space and a one dimensional smooth for time. There is a lot of choice for possible (penalized) smoothing bases, including soap film smoothers, thin plate regression splines, splines on the sphere, Gaussian Markov random fields and many others. For example using a soap film smoother for space avoids imposing correspondences between spatially adjacent areas that are in fact separated by a boundary, e.g. land. We illustrate the GAMM approach to spatio-temporal modelling using catch and effort data from the blue ling fishing industry off the northwest coast of Scotland. We will also cover how the GAMM approach relates to other methods currently used for spatio-temporal modelling of fishery catch and effort data, for example spatio-temporal models using Gaussian random fields for space combined with a first-order autoregression process for time.

Jay M. Ver Hoef,

A Review of Spatial (and Spatio-Temporal) Autoregressive Models

Conditional autoregressive (CAR) and simultaneous autoregressive (SAR) models are network-based models (also known as graphical models) specifically designed to model spatially autocorrelated data based on neighborhood relationships. I identify and discuss six different types of practical inference using CAR and SAR models, including: 1) model selection, 2) spatial regression, 3) estimation of autocorrelation, 4) estimation of other connectivity parameters, 5) spatial prediction, and 6) spatial smoothing. I compare CAR and SAR models, showing their development and connection to partial correlations. Special cases, such as the intrinsic autoregressive model (IAR), are described. CAR and SAR models depend on weight matrices, whose practical development uses neighborhood definition and row-standardization. Weight matrices can also include covariates and connectivity structures, which I emphasize, but have been rarely used. I clarify relationships between conditional (CAR) and simultaneous (SAR) autoregressive models. We review the literature on this topic and find that it is mostly incomplete. My main result is that a SAR model can be written as a unique CAR model, and while a CAR model can be written as a SAR model, it is not unique. In fact, I show how any multivariate Gaussian distribution on a finite set of points with a positive-definite covariance matrix can be written as either a CAR or a SAR model. I illustrate how to obtain any number of SAR covariance matrices from a single CAR covariance matrix by using Givens rotation matrices on a simulated example. I also discuss sparseness in the original CAR construction, and for the resulting SAR weights matrix. For a real example, I use crime data in 49 neighborhoods from Columbus, Ohio, and show that a geostatistical model optimizes the likelihood much better than typical first-order CAR models, but then use the implied weights from the geostatistical model to estimate CAR model parameters that provides the best overall optimization. Trends in harbor seals (Phoca vitulina) in southeastern Alaska from 463 polygons, some with missing data, are used to illustrate the six inference types. I develop a variety of weight matrices and CAR and SAR spatial regression models are fit using maximum likelihood and Bayesian methods. Profile likelihood graphs illustrate inference for covariance parameters. The same data set is used for both prediction and smoothing, and the relative merits of each are discussed. I show the heteroscedastic variances and correlations of a CAR model and demonstrate the effect of row-standardization. I include several take-home messages for CAR and SAR models, including 1) choosing between CAR and IAR models, 2) modeling ecological effects in the covariance matrix, 3) the appeal of spatial smoothing, and 4) how to handle isolated neighbors. I highlight several reasons why scientists will want to make use of autoregressive models, both directly and in hierarchical models, and not only in explicit spatial settings, but also for more general connectivity models. Finally, I show how CAR and SAR spatio-temporal models are network-based models with different weights for space and time. Spatio-temporal data sets are often large, and sparse matrix and other techiques could be adapted to for these data to speed computations. I consider a version of block sampling where the underlying graph can be cut so that sites are conditionally independent. This algorithm allows for parallel processing and for using vectorized calculations in R.

Toshihide Kitakado

Issues related to spatio-temporal modeling CPUE data highlighted by applications are discussed.

Maunder

Likelihood functions for including indices of abundance in stock assessment models

Indices of relative abundance derived from catch per unit effort (CPUE) are one of the main sources of data used to inform parameter estimation in fisheries stock assessment models. In contemporary statistical integrated stock assessment models, these indices are typically fit using a lognormal distribution based likelihood function as an index aggregated over all ages/sizes of fish. Any composition data is fit using an independent likelihood function, usually based on the multinomial or related distribution. In contrast, virtual Population analysis (VPA) historically fits to an index of abundance for each age. These two approaches describe two extremes in the treatment of correlations among ages in the index of relative abundance. The spatio-temporal model estimates a multivariate index of relative abundance (time and age/size) and the associated variance-covariance matrix. Preferably, this multivariate index should be fit in the stock assessment model using a multivariate likelihood function (e.g. multi-variate lognormal) using the estimated variance-covariance matrix. Unfortunately, most general stock assessment programs do not have this capability, and the index has to be either broken into separate indices for each age or into a total index and an estimate of proportion-at-age or proportion-at-length that is then treated as “compositional data”.

Length or age compositions describing the population abundance (index) are unlikely to be the same as those describing fishery catches. The catch at age/size should be calculated from the spatio-temporal model weighting the composition in each area by the catch, while the index of abundance should be calculated by weighting the composition in each area by the CPUE. The catch composition data is usually fit in the stock assessment model using a multinomial or related distribution based likelihood. However, if both the catch composition and the index of abundance composition data are used in the stock assessment then the original sampled composition data are essentially used twice. Optimally, a multivariate likelihood that includes both the catch and index data and the associated variance covariance matrix should be use. However, given the arbitrariness of weighting data and the common approach of internally estimating the weighting of composition data due to unmodeled process variation, the double use of the data is probably less of an issue than using biased composition data for indices of relative abundance. Since the composition data for the index of abundance would typically be considered the most representative and the spatio-temporal model reduces the temporal variation in selectivity caused by spatial factors, the influence of the catch composition data could be minimized by removing the catch at age directly as estimated from the spatio-temporal model paralleling a VPA or with flexible time varying selectivity if used in a contemporary statistical stock assessment model.

Contributed Abstracts

John Best

Estimation variation when fishery-dependent data is included in a spatiotemporal model

Spatially-explicit catch and effort data is increasingly available, and may provide an important source of information for estimating abundance.
There is potential to improve abundance estimates by combining fishery-independent and fishery-dependent data. To measure the potential effect of combining these data, a fishery will be simulated with both survey and fishery vessels, each with their own targeting behavior. Abundance will be estimated using only the survey observations and with both survey and fishery observations. The variation in these estimates will then be compared.
Spatially-explicit catch and effort data is likely to provide a substantial decrease in estimation variation. This should provide guidance in the expected change in estimation variation that may occur when these data are included.

Jie Cao, James Thorson, André Punt, Cody Szuwalski

Development of a size-structured spatiotemporal population model for invertebrates: individual growth, size-transitions, and natural and fishing mortality

Population dynamics and stock assessment models have historically modeled populations by tracking total abundance (often structured by age, size, and/or sex) across the entire stock. This practice implicitly assumes that individuals are equally mixed within each stratum, and overlooks the fact that marine populations are spatially patchy and locally structured. In this study, we apply spatiotemporal approach to population models by combining theory and methods from population dynamics and geostatistics. Specifically, spatiotemporal population models define population variables (density) as varying continuously across space, while estimating spatial variation as a random effect. Size-structured population models are convenient for populations where ageing information is unavailable or inaccurate, and such models are commonly used for assessment of crustacean species worldwide, e.g. crab, shrimp, and lobster. This proposed size-structured spatiotemporal model allows external specification of survival, growth, and size transition rates, and assumes that recruitment can be approximated as varying around a constant mean that varies spatially. Spatially-explicit catch data from fisheries are used for estimating annual fishing mortality across space. We use simulation approach to demonstrate that the spatiotemporal population model provides (1) accurate and precise estimates of spatial variation in population density and fishing mortality of each size class over years, (2) unbiased estimates of total abundance and model parameters, such as fishery selectivity. We apply this model to eastern Bering Sea snow crab, where spatially-explicit fishery-dependent and fishery-independent data are directly used.

Paul B. Conn, Marine Mammal Laboratory, NOAA Alaska Fisheries Science Center, Seattle, WA

Spatio-temporal models for Arctic seal abundance: successes and computational challenges

Estimation of abundance of ice-associated seals from aerial survey counts is challenging because (1) the spatial distribution of seals changes while sampling is being conducted, and (2) because the observation process is complicated by nuisance processes (e.g., incomplete detection, species misclassification). In this talk, I present some lessons from successes with modeling these data, as well as some challenges and avenues for future research. In particular, I have found that multinomial models with a single, constant abundance parameter are much stabler than models that allow total abundance to vary over time. However, such models have primarily been analyzed using a relatively complicated spatio-temporal hierarchical modeling framework with bespoken MCMC samplers, which have yielded which has led to long execution times (e.g. 8-10 days) and difficulties with conducting thorough testing. Recently, I have begun to port such models to TMB to increase execution times, but the multinomial structure does not appear to lead to sparse Hessians needed for estimation of high resolution spatio-temporal random effects.

Simon Dedman, Rick Officer, Deidre Brophy, Maurice Clark, David Reid

Automated Boosted Regression Tree software for spatial prediction of multiple life-history stock components

Boosted Regression Tree models are used to predict spatial abundance of four juvenile and mature female rays in the Irish Sea using fish survey, fishing pressure, predation and environmental variables. Model-predicted spatial abundance maps of these subsets reveal distinct nuances in species distributions with greater predictive power than maps of the whole stock.

These resulting maps are then integrated into a single easily understood map using a novel approach, standardizing and facilitating the spatial management of data-limited fish stocks.

This approach and outcomes are presented in the context of the full software suite we have written, which facilitates BRT modelling by simplifying and/or automating the majority of the processes involved.

Nicholas D. Ducharme-Barth, Kyle W. Shertzer, Robert N.M. Ahrens

ndices of abundance in the Gulf of Mexico reef fish

The Gulf of Mexico reef fish complex is socioeconomically important and is exploited by a vertical line fishery capable of high resolution spatial targeting. Indices of abundance derived from fishery dependent catch-per-unit- effort (CPUE) data are an important input to the assessment of these stocks. Traditionally, these indices have been derived from standardized logbook data, aggregated at a coarse spatial scale, and are limited to generating predictions for observed spatiotemporal strata.

Understanding how CPUE is spatially distributed, however, can help identify range contractions and avoid hyperstability or hyperdepletion, both of which can mask the true population dynamics. Vessel monitoring systems (VMS) can provide complete, high-resolution distributions of CPUE used to create abundance indices. Here we compare two methods — spatial averaging of VMS-derived catch and effort data and the result of generalized linear models applied to logbook data for generating indices, to evaluate the use of VMS-derived abundance indices in assessments of reef fish stocks. This work suggests that in fisheries where targeting occurs at very fine spatial scales, abundance indices derived from high-resolution, spatiotemporally complete data may more accurately reflect the underlying dynamics of the stock.

Francesca C. Forrestal, Michael Schirripa, C. Phillip Goodyear, Haritz Arrizabalaga, Elizabeth Babcock, Rui Coehlo, Walter Ingram, Matt Lauretta, Mauricio Oritz, Rishi Sharma, John Walter

Testing robustness of CPUE standardization and inclusion of environmental variables with simulated longline catch datasets

Increased environmental variation has caused changes in the distribution, migratory patterns, and susceptibility to various fishing gears for highly migratory marine fish. These changes become especially problematic when they manifest themselves through indices of abundance (such as catch-per-unit-effort, or CPUE) used to assess the status of fish stocks. The use of simulated CPUE data sets with known values of underlying population trends have been recommended by ICCAT to test the robustness of CPUE standardization methods. A longline CPUE data simulator (LLSIM) was developed to meet this requirement to simulate data with known properties for testing a variety of hypotheses. Effort data from the US pelagic longline fleet was paired with a volume weighted habitat suitability model for blue marlin (Makaira nigricans) to derive a simulated time series of blue marlin catch and effort from 1986-2015 with four different underlying population structures. The simulated time series were provided to stock assessment scientists to determine if the underlying population trends could accurately be detected with different methods of CPUE standardization that did and did not incorporate environmental data. While the analysts' approach to the data and the modeling structure differed, the underlying population trends were captured, some more successfully than others. However, differences in approaches highlight the importance of how variables are grouped and the criteria for including variables and factors.

Simon Hoyle

Longline CPUE indices for Indian Ocean stock assessments

In 2015 we began developing CPUE indices for Indian Ocean assessments for bigeye, yellowfin, and albacore tunas, using combined operational datasets from the Japanese, Taiwanese, Korean, and Seychelles fleets. We have addressed concerns about data quality, target change, spatial variability, and spatial and temporal variation in selectivity. I will present some of the methods used in these analyses, with a focus on spatio-temporal issues.

Peter Kuriyama

Analyzing Size data for a tuna stock

Melissa Monk

Indices of abundance for rockfish with varying knowledge of habitat

I'm looking a running indices of abundance for nearshore rockfish from onboard observer surveys of the recreational CPFV fleet. The observer program produces drift-specific catch rates, and, for north of Pt. Conception, we have high-resolution bathymetry data. Coupling the two allows us to subset data to fishing stops within the appropriate habitat. I'm going use the same data set to run indices of abundance subsetting the data using the habitat data, alpha hulls as a proxy for habitat, and using Stephens-MacCall filtering, to look at how much we gain from knowing precise habitat and how the variance estimates change.

A seasonal spatio-temporal model of summer flounder (Paralichthys dentatus) on the Northeast U.S. Shelf

Charles Perretti

The summer flounder, or fluke, (Paralichthys dentatus) is the most frequently caught recreational flatfish species on the U.S. Atlantic coast.
The recreational quota is allocated to individual states based on historical catch. However, recently, that allocation has become a point of contention due to a potential shift in the species distribution. Here we describe ongoing research to apply a seasonal spatio-temporal model to examine the distribution of summer flounder on the Northeast U.S. Shelf over the past 40 years. We use bottom trawl survey data to investigate whether the stock has shifted in recent years, and, if so, whether it can be explained by environmental variables or population structure.

Michelle Sculley

Incorporating the Spatiotemporal Distribution in the Standardization of Swordfish (Xiphias gladius) Catches in the North Pacific Ocean Hawaii-based Longline Fishery

Catch distributions of swordfish (Xiphias gladius) from the Hawaii-based longline fishery exhibit spatial heterogeneity in the North Pacific Ocean. Generalized additive models (GAMs) have been applied to explore the relative influence of spatial, temporal, and environmental predictors on the distribution of swordfish using length data from the Hawaii-based longline fishery Pacific Islands Regional Observer Program. This analysis has suggested spatial heterogeneity exists between male and female swordfish and mature and immature swordfish. Generalized linear mixed models were applied to CPUE values from the Hawaii-based longline fishery Pacific Islands Regional Observer Program catch data using spatial, temporal, and environmental predictors. This analysis has shown latitude to be an important predictor of swordfish catch as well as several environmental variables. Currently, work to standardize CPUEs using the R package VAST has successfully replicated the north to south heterogeneity observed in the GLM standardization, but there are several challenges to using VAST when incorporating catchability covariates and accounting for the strong seasonality in the fishery.

M.S.M. Siddeek, J. Runnebaum, C. Siddon, J. Zheng, B. Daly, and W.B. Gaeuman

Developing abundance indices for Aleutian Islands Golden King Crab using fisheries-dependent and fisheries-independent data in a spatio-temporal model

The Aleutian Islands golden king crab (*Lithodes aequispinus*) stock assessment moved from a Tier 5 (non-model-based) to a Tier 3 (model-based) approach in 2017 after the development of a length-structured model that includes the use of a standardized index of abundance. Catch-per-unit-effort (CPUE) was standardized with a generalized linear model (GLM) to remove biases of captain, area, and soak time. CPUE data were collected by onboard crab observers from pots sampled from longline strings of up to 40 or more pots. Observers have a 7 pots/day quota in those fisheries. There can be more than one pot sampled from a single string, depending on the number and size of strings hauled on a day. To obtain a more fishery-independent abundance index, Alaska Department of Fish and Game has completed the first three years (2015–2017) of an agency/industry cooperative golden king crab survey in the eastern Aleutian Islands. Survey pots were likewise sampled from longline pot strings, but strings were fished in areas determined by the survey design rather than normal fishing operations. Given the limited extent of the resulting time series of survey data, we propose utilizing a spatio-temporal delta generalized linear mixed model to develop separate sets of abundance indices based on the fishery-dependent observer data and on the fishery independent survey data. This approach requires that several key questions be resolved: 1) how to quantify CPUE (e.g., pots vs strings); 2) how to evaluate the concept of “area swept” for pot fisheries within the spatio-temporal model; 3) how to incorporate a model selection procedure to select the appropriate covariates for capturing differences in fishing effect; 4) how to account for a decrease in vessel number and fishing area since post rationalization (i.e., after the 2004/05 fishery); and 5) how to correct for potential bias due to the use of fisheries-dependent data.

Brian Stock, Eric Ward, Brice Semmens

Random forests for spatiotemporal modeling of fisheries data

Recent simulation tests have shown that random forests (RF) outperform semi-parametric models, such as generalized additive models (GAMs) and Gaussian Markov random fields (GMRFs), at predicting spatiotemporal fisheries catch. Can RF be used to develop indices of abundance with uncertainty? Can multivariate RF integrate length-frequency data and predict multispecies density? We demonstrate how RF can be used to derive an index of abundance for an example yellowfin tuna CPUE dataset. We hope to stimulate discussion by highlighting the relative merits of RF versus other methods.

Josymar Torrejón – Magallanes, Wencheng Lau – Medrano, Daniel Grados, Gladys Castillo and Ana Medina

Spatio-temporal distribution modeling and abundance index of perico (Coryphaena hippurus) in the Pacific Ocean off Peru

Perico, also known as dolphinfish, mahi mahi or dorado, is an epipelagic higly migratory species, mainly oceanic, distributed worldwide in tropical and subtropical waters. In Peru, the fishery of perico is carried out by the longline artisanal fleet and currently their landings represents more than 50% (~ 50 000 tons) of the total landing of perico around the world. The aim of this study are modelling abundance/availability trends and the spatio-temporal distribution of perico in the Pacific Ocean off Peru during the 2009 – 2017 period as a function of spatial (longitud, latitude), temporal (year, month) and environmental variables (sea surface temperatura, salinity and chorophyll). For model building we considered the spatial/temporal autocorrelation and used two approaches: The delta Generalized Additive Models (delta - GAM) and the spatial-temporal zero inflated Bayeasian model (ZIB). A detailed discussion is made using the main variables that trigger spatial and temporal changes in distribution and abundance/availability trend. Spatio-temporal analysis gives us clues about habitat preferences of perico (hotspots) in the transitional areas of Peru (coastal and oceanic), likewise abundance/availability trend would help to develop conservation management.

Laura Tremblay-Boyer, Samuel McKechnie, Graham Pilling, John Hampton

Accounting for the effects of oceanography on catchability and recruitment in basin-wide standardized indices of abundance for Pacific yellowfin and bigeye tuna

Fishery-dependent data are the main source of relative abundance trends informing stock assessments for bigeye, yellowfin and albacore tuna in the Western and Central Pacific, despite well-documented issues like changes in targeting and gear configuration since the 1960s. We present here preliminary results for the Pacific-wide standardization of bigeye and yellowfin logsheet catch-per-unit-effort for the Japan, Korea and Chinese Tapei longline fleet from 1960 to 2015 based on a spatiotemporal generalized linear model. In parallel, we discuss potential approaches aiming to account for the non-random distribution of fishing effort over space and time in our study region. This key feature of fishery-dependent datasets (fishing effort increasing in locations with higher catch rates) violates a basic assumption of geostatistical approaches, which have so far mostly been applied to fishery-independent datasets.

We account for a non-linear effect of thermocline depth and sea surface temperature on catchability and recruitment, as the El Nino-Southern Oscillation can impact spatial and temporal patterns in catch rates for tuna by both modifying the available vertical habitat and improving spawning conditions. Given the extent of the dataset (~11 million records), we also performed subsampling to improve the computational performance of the model, and we review the impact of various subsampling strategies on the final abundance indices. Finally, we developed new diagnostic tools to improve the detection of spatial and temporal trends in model residuals and demonstrate these briefly.

Katyana A. Vert-pre, James T. Thorson, Thomas Trancart, Eric Feunteun

Assessing Functional Structure in European Marine Ecosystems Using a Vector-Autoregressive Spatio-Temporal Model.

In marine ecosystems, spatial and temporal species structure is an important component of ecosystems' response to anthropological and environmental factors. Although spatial distribution patterns and fish temporal series of abundance have been studied in the past, little research has been allocated to the joint dynamic spatio-temporal functional patterns in marine ecosystems and their use in multispecies management and conservation. Each species represents a function to the ecosystem, and the distribution of these species might not be random. A heterogeneous functional distribution will lead to a more resilient ecosystem to external factors. Applying a Vector-Autoregressive Spatio-Temporal (VAST) model for count data we estimate the spatiotemporal distribution, shift in time, and abundance of 140 species of the Eastern English Chanel, Bay of Biscay and Mediterranean Sea using historical Data Collection Framework series. From the model outputs, we determined spatio-temporal clusters, calculating p-values for hierarchical clustering via multiscale bootstrap resampling. Then we designed a functional map given the defined cluster. We found that the species distribution within the ecosystem had a clustered structure. Indeed, species evolved in space and time in clusters. Moreover, these clusters remained similar over time deriving from the fact that species of a same cluster often shifted in sync, keeping the overall structure of the ecosystem similar overtime. Knowing the co-existing species within these clusters could help with predicting data poor species distribution and abundance. Further analysis is being performed to assess the ecological functions represented in each cluster.

John Walter, Arnaud Gruss, Elizabeth Babcock, Matt Lauretta, Francesca Forrestal, Michael Schirripa, Clay Porch

Evaluation of three standardization methods to estimate CPUE from observer data

In practice, observer data is almost always a subset of the entire fishery and often rarely a random subset due to the vagaries of allocating observers to vessels and the potential changes in fishing practices due to having an observer onboard. Observer data also represents multiple, correlated observations from individual trips, rather than independent sets. This makes observer data potentially more challenging to develop standardized CPUEs and potentially even more reliant upon spatial estimation methods proposed by Campbell (2015) and Thorson (2015). We apply three methods: standard delta-lognormal approach, Campbells (2015) spatial gap filling method and Thorson's geostatistical estimator (VAST) to known data produced from the longline catch-per-unit-effort (CPUE) data simulator LLSIM (Forrestal et al., 2017; Goodyear et al., 2017). The LLsim data is split into three different subsets; the full dataset, a random subset of 10% of the trips and a biased subset of 10% of the trips where trips were selected with a bias against catch of blue marlin. This bias was intended to mimic an 'observer effect' where the placement of an observer might alter fisher's behavior to avoid areas/time/fishing methods that might have higher bycatch. We then compare performance of the three methods across the three dataset constructions.

Haikun Xu

The spatiotemporal dynamics of yellowfin tuna (Thunnus albacares) in the Eastern Pacific Ocean

We apply a size-structured, spatiotemporal, delta-generalized linear mixed model to the size-structured catch rate of yellowfin tuna (Thunnus albacares) for the dolphin-associated purse-seine fishery in the Eastern Pacific Ocean. The spatiotemporal model is able to standardize the indices of abundance for each size bin as well as the entire population as a whole, which are then compared to the corresponding standardized indices based on conventional methods such as generalized linear model and generalized additive model. The spatiotemporal model may outperform those conventional methods because it accounts for spatial and temporal autocorrelations in the residuals of encounter probability and positive catch rate, and hence, allows for estimating the density in unsampled areas via imputation. Also, we use the spatiotemporal model to predict the spatial distribution of yellowfin tuna for each size bin. In this way, we can examine the difference in the spatial distribution for various life stages, namely, the condition of size-segregation of available habitat. Lastly, we examine whether and how the spatial distribution of yellowtail tuna is affected by environmental conditions (e.g., sea surface temperature) at each life stage. To do so, we incorporate the potentially important environmental drivers as covariates in the spatiotemporal model and compare the marginal standard deviation of spatial and spatiotemporal variations with and without those environmental conditions.

Focus questions

Statistical issues

How to deal with barriers?

There are many instances where the area occupied by a stock is not simple due to a variety of barriers, the most obvious being a land mass such as a peninsular. The barrier will alter how the correlation function should be formulated. How can these be taken into consideration?

What spatial correlation function should be used?

The Matern function is commonly used because it is computationally efficient. Are there alternatives? Is there a way to test between different correlation functions?

Can the correlation function change spatially?

Some analyses cover large areas of ocean with different oceanographic features causing the spatial correlation to differ. Are there approaches that allow the correlation function to vary spatially. For example, can the correlation function be based on covariates?

What temporal correlation function should be used?

The first order autoregressive model is often used for the temporal component. Is this function flexible enough or are others more appropriate? Is there a way to test between different correlation functions?

Should geometric anisotropy be used?

Changes in CPUE may be different latitudinally compared to the changes longitudinally, in these cases the parameters of the correlation function should be estimated separately for latitude and longitude (*e.g.*, termed geometric anisotropy). Does modelling geometric anisotropy greatly increase the computational demands? What tests can be used to determine if geometric anisotropy should be used?

How should multiple fisheries with different spatial distributions and selectivities be combined?

Commonly, a single fishery does not cover the whole spatial distribution of a stock and multiple fisheries may have to be combined to get a full picture of the spatial distribution and produce a comprehensive index of abundance. However, different fisheries may have different catchability and different age/size selectivity. Therefore, care need to be taken when combining data from different fisheries. Intuitively, if there is some spatial overlap, it might be possible to estimate the differences in selectivity and catchability.

Computational issues

How should the data be aggregated?

CPUE data is often available by set of the gear and there may be numerous sets Individual sets in each location. However, including each individual set in the analysis may make the analysis too computationally intensive. Could the data be aggregated by spatial strata and how would the variance among sets be accounted for? The aggregation would be more complicated as covariates are added to the analysis and the data would have to be aggregated by those categories as well. Use of continuous covariates may invalidate the use of aggregation and there may be a tradeoff between including a covariate and computational efficiencies obtained by aggregation.

What is the appropriate spatial resolution?

There is a trade-off between the spatial resolution of the analysis and the computational demands. The spatial analysis can be conducted at a finer resolution than the data and then the model predictions aggregated to fit to the data. However, it is not clear if there are advantages to this or if it causes convergence issues. The analysis can also be conducted at a courser resolution than the data and the data aggregated to fit the model. This will reduce computational demands, but may degrade the results.

CPUE Issues

How to determine edges when calculating the index of abundance

The index of abundance is calculated by summing up the densities in each spatial strata. As spatial strata are added by expanding the area covered by the analysis, the index may change. Spatial strata on the edge of the area may have few or no data points and may be heavily influenced by their neighbors. It is not clear whether these strata should be included in the analysis, particularly since it may not be known if the effort in these areas is low because the densities are low or for another reason.

What factors should be considered related to density and what related to catchability?

The index of abundance should only be constructed using the factors related to density. Therefore, any factors used in the model should be separated into catchability and density. In general, the space and time factors are considered related to density. However, these may be correlated with factors related to catchability. For example, habitat type might influence catchability and may change spatially. Therefore, habitat should be considered as a catchability factor and not included in the index calculation. On the other hand, a species might be more abundant in an area because of the habitat type. So, it may be difficult to attribute a factor to either density or catchability.

Is there a method to reduce the bias of targeting that creates nonrandom sampling in CPUE data?

Targeting typically means that areas of high abundance are sampled (fished) more proportionally than areas of low abundance than would be expected by simple abundance weighted sampling. This may bias the results. What is the bias and how can it be avoided? Can inclusion of other species in the analysis be used to define which sets are targeted and, if so, how can this information be used in the analysis be used to reduce the bias. Both target and non-targeting sets may be needed to get the spatial coverage required for the analysis.

Vertical habitat

Most species have different preferences for depth and other oceanographic conditions. Therefore, it may be important to take the third dimension, depth, into consideration in the spatio-temporal analyses. How do the approaches need to be modified and what effect will it have on the computational demands.

Composition data

How should the age/size composition data be treated in the spatio-temporal model?

The simplest approach is to calculate the CPUE by age/size and simply fit to that data using a log-normal distribution based likelihood function or something more sophisticated (e.g. a zero-inflated distribution based likelihood). However, often the composition data is sampled at a lower rate than the CPUE data. Therefore, it may be more appropriate to fit to the aggregated CPUE data separately from the composition data, similar to how this is modelled in contemporary statistical stock assessment models. For example, a lognormal distribution based likelihood function could be used for the aggregated CPUE and a multinomial distribution based likelihood used for the composition data. Similar to data weighting issues in stock assessment, care needs to be taken when defining and estimating the variance parameters of each data type.

Different resolution between composition and catch/effort

Size composition data is often at a different (usually coarser) spatial resolution than catch and effort data. The model could still be defined on the finest resolution and the data fit on the resolution for which it is observed after aggregating the model predictions. Intuitively, the covariance function that smooths over space should allow for the courser resolution of the data. However, it is not clear if there would be convergence issues. The CPUE and composition data would have to be fit using different likelihood functions as discussed in the previous focus question.

What age/length correlation function should be used?

The first order autoregressive model has been used for the age/length component. Is this function flexible enough or are others more appropriate? Is there a way to test between different correlation functions?

How common is the existence of a difference between the age/size structure of the population as represented by the index of abundance and the catch?

The main reason behind including the age/size data in the spatio-temporal model is to deal with the differences in composition between the stock and the catch due to spatial structuring of the age/size of fish and differences in the spatial distribution of the catch relative to the population. How common this is will determine the need to develop these methods.

Other Methods

What are the advantages of the other methods?

There are a variety of methods that can be used to model spatio-temporal data. We have focused on Generalized linear mixed models with spatio-temporal random effects implemented using a Gaussian random field (GRF), which is a computationally efficient approach for implementing multi-dimensional smoothers. Other methods include 3-D or 4-D general additive models, soap film smoothers, general additive mixed effect models, regression trees, and neural networks. The different methods have their own advantages and disadvantages. For example, soap film smoothers have been used to deal with boundaries in space. What are the advantages and disadvantages of each method and when are they the most appropriate to use?

How can an index of abundance be created from a regression tree?

Regression tress are typically used for categorizing data or making predictions. Creating an index of abundance is a different objective and methods need to be developed.

Stock Assessment Issues

How should indices of abundance from the spatio-temporal analysis be included in the stock assessment model?

The ultimate goal of the analysis is to produce CPUE based indices of relative abundance to use in the stock assessment. The simplest approach is to generate age/size specific indices and fit them independently in the stock assessment, which is similar to what has historically been done in Virtual Population Analysis (VPA). However, this ignores the correlation among the age/length groups. A more statistically rigorous approach would be to produce a multivariate index with associated covariance matric and fit this in the stock assessment model using a multivariate likelihood function. Unfortunately, most general stock assessment programs do not have the capacity to use multivariate likleihoiod fucntions for indices of abundance. Therefore, the index has to be either broken into separate indices for each age or into a total index and an estimate of proportion-at-age or proportion-at-length that is then treated as “compositional data”.

How should the composition data for the catch be included in the stock assessment?

The composition data is used both for the indices of abundance and the catch. However, the composition data will be weighted across space differently, by CPUE and catch, respectively. The calculation will generally mean that the data will be used twice in the stock assessment model, even though it is weighted acrosss space differently, which is a violation of standard statistical practices. Ideally, a multivariate distribution of realtive abundance and catch by age/size should be developed in the spatio-temporal model and used in the assessment, but this is probably unrealistic. Given the arbitrariness of weighting data and the common approach of internally estimating the weighting of composition data, the double use of the data is probably less of an issue than using biased composition data for indices of relative abundance. Since the composition associated with the abundance index represents the population, only factors representing the density effects on the composition data should be used in calculating the age/length structure of the index, and catchability effects should be ignored. However, when calculating the fishery catch age/length structure both the density and catchability effects should be included.