2012

Geostatistical spatiotemporal modelling with application to the western king prawn of the Shark Bay managed prawn fishery

Ainslie M. Denham

Edith Cowan University

Copyright Warning

EDITH COWAN UNIVERSITY
FACULTY OF COMPUTING, HEALTH AND SCIENCE
SCHOOL OF ENGINEERING

GEOSTATISTICAL SPATIOTEMPORAL MODELLING WITH APPLICATION TO THE WESTERN KING PRAWN OF THE SHARK BAY MANAGED PRAWN FISHERY

AINSLIE DENHAM

PRINCIPAL SUPERVISOR:
ASSOCIATE PROFESSOR UTE MUELLER

APRIL 2012

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

ABSTRACT

Geostatistical methodology has been employed in the modelling of spatiotemporal data from various scientific fields by viewing the data as realisations of space-time random functions. Traditional geostatistics aims to model the spatial variability of a process so, in order to incorporate a time dimension into a geostatistical model, the fundamental differences between the space and time dimensions must be acknowledged and addressed. The main conceptual viewpoint of geostatistical spatiotemporal modelling identified within the literature views the process as a single random function model utilising a joint space-time covariance function to model the spatiotemporal continuity. Geostatistical space-time modelling has been primarily data driven, resulting in models that are suited to the data under investigation, usually survey data involving fixed locations.

Space-time geostatistical modelling of fish stocks within the fishing season is limited as the collection of fishery-independent survey data for the spatiotemporal sampling design is often costly or impractical. However, fishery-dependent commercial catch and effort data, throughout each season, are available for many fisheries as part of the ongoing monitoring program to support their stock assessment and fishery management. An example of such data is prawn catch and effort data from the Shark Bay managed prawn fishery in Western Australia. The data are densely informed in both the spatial and temporal dimensions and cover a range of locations at each time instant. Both catch and effort variables display an obvious spatiotemporal continuity across the fishing region and throughout the season. There is detailed spatial and temporal resolution as skippers record their daily fishing shots with associated latitudinal and longitudinal positions. In order to facilitate the ongoing management of the fishery, an understanding of the spatiotemporal dynamics of various prawn species within season is necessary. A suitable spatiotemporal model is required in order to effectively capture the joint space-time dependence of the prawn data.

An exhaustive literature search suggests that this is the first application of geostatistical space-time modelling to commercial fishery data, with the development and evaluation of an integrated space-time geostatistical model that caters for the commercial logbook prawn catch and effort data for the Shark Bay fishery. The model developed in this study utilises the global temporal trend observed in the data to standardise the catch rates. Geostatistical spatiotemporal

variogram modelling was shown to accurately represent the spatiotemporal continuity of the catch data, and was used to predict and simulate catch rates at unsampled locations and future time instants in a season. In addition, fishery-independent survey data were used to help improve the performance of catch rate estimates.

DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education;

(ii) contain any material previously published or written by another person except where due reference is made in the text of this thesis; or

(iii) contain any defamatory material.

Signed: _____ _____ Date: 22/03/2012

ACKNOWLEDGEMENTS

I would like to thank the Department of Fisheries (WA) for allowing me to use their logbook catch and effort data for my research, and to fisheries scientists and support staff at the Western Australian Fisheries and Marine Research Laboratories for their advice and assistance.

To my husband Tobe and my two beautiful children Britney and Taj, my heartfelt thanks for their understanding and patience throughout the duration of my studies. Their physical, emotional and financial support was never-ending and gave me the ability to complete my thesis. I am also thankful to my parents and to my sister Kylie for their support and encouragement through my studies. They are a constant source of love and strength in my life and I cherish them all so much for their guidance and motivation. Thanks also to other members of my extended family and close friends, who were very supportive throughout my doctoral studies and who helped out with babysitting of my children.

Finally, I would like to thank my principal supervisor Associate Professor Ute Mueller for her endless supply of gummy bears and home-made chocolate covered orange pieces, her support and encouragement throughout my doctoral studies at Edith Cowan University, and without whom this journey, and even more importantly its completion, would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PART I.

GENERAL INTRODUCTION, THEORETICAL FRAMEWORK, DATA DESCRIPTION AND METHODOLOGY

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND AND SIGNIFICANCE

Geostatistics is a class of techniques traditionally used to analyse and predict values of a spatially distributed variable, which are implicitly assumed to be spatially correlated. The prefix geo- comes from geology, since geostatistics has its origins in mining (Matheron, 1971; Journel & Huigbregts, 1978). Several books have been written on the subject of geostatistics, ranging from those which mainly focus on the applications of geostatistics (Isaaks & Srivastava, 1989; Goovaerts, 1997; Deutsch & Journel, 1998) to ones that treat geostatistics from a more mathematical and statistical point of view (Cressie, 1993; Chiles & Delfiner, 1999; Stein, 1999). With its roots in mining, the focus of geostatistics was with static spatial variables, as the target variables are often constant for time periods much smaller than the geologic time scale. Over the past few decades, the usefulness of geostatistics has been extended from mining to a wider class of environments including the land, atmosphere and ocean. Many of these fields involve variables that exhibit not only spatial correlation but temporal correlation. Consequently, geostatistics has been extended to include spatiotemporal modelling to address variables that change in space and time. The joint analysis of space and time is based on the same philosophy as the spatial analyses.

The extension of geostatistical techniques into the space-time domain is not straightforward. From a mathematical perspective, the domain of a spatiotemporal process may appear to have simply incorporated one more dimension, but this does not recognise that space and time are completely different physical notions. For example, characteristics of the temporal processes that cause temporal variation, such as periodicity associated with daily or seasonal patterns, are often known to some degree, Such periodic variation is less common in space. Therefore, a realistic statistical model characterising the spatiotemporal behaviour of a variable has to take into account the differences between variation in space and in time. Although the inference and prediction techniques remain the same, the main difference relates to the modelling of the second-order characteristics of the process. The fundamental difference between space and time must be acknowledged through the covariance function (Gneiting, 2002).

Geostatistical spatiotemporal models provide a probabilistic framework for data analysis and predictions that builds on the joint spatial and temporal dependence between the observations (Kyriakidis & Journel, 1999a). Although deterministic models involving the appropriate governing equations can more readily account for the physical processes of the phenomenon under study, these models are typically not easily known. Moreover, these models typically require a large number of input parameters which are difficult to determine (Armstrong, 1998). Stochastic models are usually based on a small number of parameters that can be modelled and are aimed at building a process that mimics some of the behaviour of the observed spatiotemporal process without knowing the underlying governing equations. The stochastic model can be regarded as an alternative approach when a more elaborate, physically-based model is unavailable. Geostatistics focuses on making a description of what has been observed by capturing the main structural features from the data. The randomness of the stochastic process is a measure of the uncertainty of the process. Space-time geostatistical modelling is an empirical approach where the model is specified and its parameters estimated from observed data. Knowledge of the subject matter is integrated in the modelling process. A stochastic model of the space-time variable of interest is postulated, from which the space-time covariance can be derived. Although not physically based, this model must appreciate the fundamental differences between spatial and temporal variation and must include these differences in its structure and parameterisation (Kyriakidis & Journel, 1999a; Heuvelink & Griffith, 2010)

Geostatistical spatiotemporal modelling considers the data as realisations of space-time random functions. A review article (Kyriakidis & Journel, 1999a), concerning the geostatistical space-time models that had been used or suggested in the literature until that time, identified two conceptual viewpoints. These two approaches still dominate the field of geostatistical spatiotemporal modelling. The first involves a single spatiotemporal random function model based on a joint space-time covariance function to model the spatiotemporal continuity. This is the most favoured approach for current studies (Gething et al., 2008; Denham & Mueller, 2009; Garcia-Cabrejo et al., 2009; Russo et al., 2009; De Iaco, 2010; Heuvelink & Griffith, 2010). The second approach views the spatiotemporal process as either a set of temporally correlated random functions (Egbert & Lettenmaier, 1986; Goovaerts & Sonnet, 1993; Papritz & Flühler, 1994; Bogaert & Christakos, 1997) or a set of spatially correlated time series (Solow &

Gorelick, 1986; Rouhani & Wackernagel, 1990; Rouhani et al., 1992). The spatiotemporally continuity of either of these multiple models is modelled via a linear model of coregionalisation.

Whilst estimation or simulation within a spatiotemporal geostatistical framework proceeds in much the same way as it does in a purely spatial setting, the greatest challenge lies in the definition of a realistic stochastic model that is deemed to imitate the behaviour of the given data and in characterising and estimating the space-time correlation of that model. The stochastic model defines the given data as a particular (partial) realisation of a random function. Statistical analysis is based on this single realisation which may appear contradictory to the classical approach of statistics. This situation complicates the inferential process but is reasonable with the use of some assumptions for the model.

Stationarity is one of the simplifying assumptions of the process under study in spatial geostatistics in order to estimate the spatial covariance. Similarly, it is used in spatiotemporal geostatistics to estimate the spatiotemporal covariance function (Gneiting et al., 2007; Bruno et al., 2009). Strict stationarity assumes that the statistical properties of the space-time process do not change over time or between locations. This means that any pair of spatiotemporal random variables has the same multivariate cumulative density function, whatever the translated vector between them. Spatiotemporal stationarity implies both spatial and temporal stationarity. It is a very restrictive property and one that is difficult to test. A less restrictive property called second-order stationarity is introduced that only considers the first two moments of the spatiotemporal random function. A random function is said to be second-order stationary if its expected value (mean) is constant over the area and its space-time covariance function depends only on the spatial and temporal separation of points. The absolute spatiotemporal coordinate does not affect the covariance as it only depends on the spatiotemporal separation. A milder hypothesis is to assume that for every translation vector the increment is second order stationary. This form of stationarity is known as intrinsic stationarity, and requires only that the expectation and variance of the increments should exist and, again, that they should depend only on the spatiotemporal separation, not on the locations themselves.

If none of the stationarity assumptions hold then we have the more general case of a non-stationary random function. Non-stationarity is a common feature of many spatiotemporal processes, in particular those observed in the earth sciences (Schabenberger & Gotway, 2005). Sources of non-stationarity may be either a non-constant mean, non-constant variance or spatially and/or temporally varying covariance function. Changes in the mean function can be

accommodated in spatiotemporal models by expressing the mean function in terms of the spatiotemporal coordinates or other related variables, whilst the variance can be stabilised by transformation of the response variable. By removing these sources of non-stationarity, the resulting variable can often be analysed via a (second-order) stationary spatiotemporal random function model. When the covariance function varies over space and/or time, the ability to inspect the second-order structure by considering only the distances between the various spatiotemporal locations is no longer possible and requires advanced techniques to reduce it to a model that employs a stationary covariance structure (Haslett & Raftery, 1989; Sampson & Guttorp, 1992; Elmatzoglou, 2006).

For either of the two spatiotemporal geostatistical approaches outlined previously the model can be decomposed into the sum of a trend and residual component. Some aspects of spatial and/or temporal non-stationarity of the mean function can be included within the trend component. The remaining (residual) component is assumed to be stationary and its variability is modelled by a stationary covariance/semivariogram function. The trend component can be viewed as deterministic or stochastic. A deterministic trend function of the spatiotemporal coordinates is adopted in many studies (Armstrong et al., 1993; Heuvelink et al., 1997; De Cesare et al., 2001a; De Iaco et al., 2001). Alternatively, a stochastic mean models the space-time distribution of the mean process as a non-stationary spatiotemporal random function (Kyriakidis & Journel, 1999a). The spatiotemporal deterministic or stochastic mean can also be decomposed into spatial and/or temporal components. These components can be considered to be additive or multiplicative and in some instances the trend component may be considered to be purely spatial or purely temporal.

The next issue in the single spatiotemporal random function model approach is how to specify and model the space-time covariance structure. The fitting of a space-time covariance model with simple or closed-form expression is essential. One approach to covariance modelling, known as the metric model, is to simply consider time as another dimension which requires an appropriate metric in space-time (Dimitrakopoulos & Luo, 1994). This has been used in several studies (Bilonick, 1985; Figueira et al., 1999; Snepvangers et al., 2003; Jost et al., 2005) and typically involves a mixture of geometric and zonal anisotropies. Geometric/zonal anisotropy involves varying ranges/sills in the directions of maximum and minimum continuity. However, the modelling of two spatial dimensions and one temporal dimension is not analogous to that involving three spatial dimensions as there are fundamental differences between the

spatial and temporal dimensions. Space and time have different scales and distance units that are physically incomparable and in contrast to the temporal domain there is no clear ordering of data in the spatial domain. A joint space-time metric does not account for these differences so other ways were considered to model the spatiotemporal covariance.

Separable models for the covariance function aim to decompose the spatiotemporal covariance into a suitable combination of a purely spatial covariance and a purely temporal covariance. The linear model involves a sum of spatial and temporal covariances (Rouhani & Hall, 1989) which correspond to a zonal anisotropy model as outlined previously. The product model is an alternative separable model involving the product of a spatial covariance and a temporal covariance (Haas, 1995; De Cesare et al., 1997; Lophaven et al., 2006). The use of separable spatiotemporal covariance models greatly reduces the number of parameters that need to be estimated and thus makes estimation easier. This structure is often assumed more for practical than for substantive reasons and implies that the spatial structure is the same at all time points and the temporal structure is the same at all locations (Cressie & Huang, 1999). The separable covariance models considered by these researchers are limited in that the variability in space and time do not interact. Such models correspond to the assumption of two separate processes (one temporal and one spatial) acting independently from each other. They correspond to a simplistic random function that will often fail to model a physical process adequately.

Non-separable spatiotemporal stationary covariance models aim to capture the joint space-time variability of spatiotemporal data. The product-sum model is a non-separable covariance model that combines sums and products of the spatial and temporal covariance models (De Cesare et al., 2001b). Other classes of valid space-time covariance structures,which employ advanced methods, have been proposed (Jones & Zhang, 1997; Cressie & Huang, 1999; De Iaco et al., 2001; Gneiting, 2002; Ma, 2002, 2003; Kolovos et al., 2004; Stein, 2005; Ma, 2008; Mateu et al., 2008). There has also been some development of non-separable spatiotemporal models which are not limited to stationary data (Christakos & Hristopulos, 1998; Christakos, 2000; Fernandez-Casal et al., 2003) but there are few examples of the practical implementation of these non-stationary covariance models.

The approach of temporally correlated random functions or spatially correlated time series to geostatistical spatiotemporal modelling arose as there often exists an imbalance between the quantities of temporal and spatial data. A data set where the sampling is denser in

space than in time can be viewed as a realisation of a set of temporally correlated (spatial) random functions (Goovaerts & Sonnet, 1993). A linear model of coregionalisation is then fitted to the simple and cross covariances of the (spatial) random functions associated with each time instant, which does not require stationarity in time. No temporal interpolation or extrapolation is possible. The computation and modelling of the associated covariances become rather cumbersome as the number of (spatial) random functions increases. Spatial non-stationarity can be handled by decomposing the (spatial) random functions into a non-stationary deterministic or stochastic trend component and a stationary residual component. A linear model of coregionalisation is fitted to the stationary residual random functions.

In a similar manner, a data set that is more densely sampled in time than in space can be modelled via a set of spatially correlated time series (Rouhani & Hall, 1989). This requires fixed spatial locations at each time instant. The analysis proceeds as outlined for a set of temporally correlated random functions. As before, the computation and modelling of the associated covariances become cumbersome as the number of time series increases. The time series at each location can be decomposed into a non-stationary deterministic or stochastic trend component and a stationary residual component to address temporal non-stationarity. The residual time series is fitted with a linear model of coregionalisation. In the basic form of this approach, no spatial interpolation or extrapolation is possible. However, it is possible to extend this approach to a continuous spatial domain by determining temporal covariance models or time series independently at each fixed location and then regionalising these location-specific parameters in space (Kyriakidis & Journel, 1999b, 2001b, 2001a). This allows temporal covariance models or time series to be determined at unsampled locations and alleviates the computational effort associated with the number of (temporal) covariances. This time series technique may also be employed to model the trend component of a single spatiotemporal random function model (De Cesare et al., 1997; Heuvelink et al., 1997).

An example of spatiotemporal data arises in the Shark Bay prawn fishery, which is located within the waters of Shark Bay off the mid west coast of Western Australia. Commercial fishing logbook data for various prawn species are available detailing the series of trawl shots for each vessel with the starting location for each trawl shot along with the relevant catch and effort information. Western king prawns (*Penaeus latisulcatus*) are the dominant species, comprising about 70% of the catch, with brown tiger prawns (*Penaeus esculentus*) chiefly making up the remaining 30%.

Tiger prawns can live for over two years and king prawns can live for up to four years, although animals greater than two years are rarely caught under current harvesting practices. King prawns have the ability to spawn numerous times throughout the year producing approximately 100000 to 700000 eggs per spawning. Tiger prawn spawning levels are lower at approximately 50000 to 400000 eggs per spawning. Prawns mature at about 10-12 months of age and spawning occurs in offshore waters. The larvae drift shoreward to shallow water where they remain in the nursery grounds for three to six months while maturing. At this point they migrate offshore and enter the trawl fishing grounds. This migration takes place in the summer and autumn of each year and is termed recruitment to the fishery. King and tiger prawns generally live in coastal waters down to a maximum depth of about 80m and are usually caught by trawlers over hard, sand sediment substrates. (Kangas et al., 2006)

The Shark Bay prawn fishery is the largest prawn fishery in Western Australia. It is an otter-trawl fishery that began in 1962. The fishery is “input controlled” with a complex series of management restrictions, including limited entry, boat size and gear controls, and spatial and temporal closures. The opening/closing dates vary each year depending on environmental conditions, moon phase and the results of recruitment surveys. Large areas of the fishery, particularly inshore, are closed to trawling to protect sensitive habitats and juvenile prawns. Intra-seasonal closures are implemented in the fishery to reflect the recruitment pattern of prawns in Shark Bay. In addition, the Department of Fisheries will open up additional areas throughout the season, which allows the fishing effort to be spread over a wider area as well as between species. (Kangas et al., 2006)

Management controls for this fishery are based on the commercial catch and effort data (Sporer & Kangas, 2004). Traditionally, modelling of Shark Bay prawn logbook data has been in the form of time series analyses involving aggregated daily, monthly or annual catches over the whole fishery or (relatively large) fishing areas (Penn & Caputi, 1985; Penn et al., 1995; Hall & Watson, 2000; Harman, 2001) without taking detailed spatial variation into account. These are not simply empirical descriptions of the time series of data but attempt to capture aspects of the relevant biology and fishery processes. When the catch and effort data of a fishery are summarised over large areas and times, the spatial and temporal resolution that can be useful for understanding patterns and formulating management decisions in fisheries is removed (Telfer, 2010). The majority of traditional fisheries research focuses on the temporal variability of the fishery population. However, an understanding of the spatial variability in fisheries can

also provide important information to management when assessing the fish population and the resources upon which it is based. By incorporating both temporal and spatial information, a spatiotemporal model for the king and tiger prawn distribution in the Shark Bay prawn fishery could be a useful addition to the set of assessment tools used for the ongoing management of the fishery.

Geostatistics was first applied for fisheries management by estimating the spatial distribution of shellfish stocks (Conan 1985) and is a useful step in fisheries science to address the need to investigate spatial patterns in abundance and provide fishery managers with another tool for exploring the distribution of a fish population. Several studies have used spatial geostatistics to study the distribution of aquatic stocks using survey data in order to characterise the spatial continuity of a measure of an aquatic species (Fletcher & Sumner, 1999; Kleisner et al., 2010b), to estimate catches and their variances (Fernandes & Rivoirard, 1999), to optimise sampling strategies (Petitgas, 1996) and to map the estimated spatial distribution of a population (Maravelias et al., 1996; Castrejón et al., 2005; Rufino et al., 2005; Adams et al., 2008; Addis et al., 2009; Murenu et al., 2010) by taking into account the spatial structure of the population. These studies use fishery-independent survey data obtained primarily for the research or as part of an ongoing monitoring program for the fishery. Analysis is limited to the spatial aspect by considering data for a given fishing period or season where the temporal component can be ignored, or by comparing independent spatial analyses for various fishing periods or seasons (Faraj & Bez, 2007; Colloca et al., 2009; Kleisner et al., 2010a).

Fishery management is based on the estimation of the abundance and sustainability status of fish stocks, which requires characterisation of the population of the target species. The stock assessment of a fish stock provides decision makers with much of the information necessary to make reasoned choices. It describes the past and current status of a fish stock and also attempts to make predictions about how the stock will respond to current and future management options. The Shark Bay fishery is required to report annually to the government and community on the status of the fishery, given that it is utilising a community resource. The management aims of the Shark Bay fishery are related to the protection and management of the fishery and include sustaining prawn stocks and maintaining economic viability. The primary operational objective is to maintain the spawning stock of each prawn species at or above a level that minimises the risk of recruitment overfishing (Kangas et al., 2006).

The assessment of the status of a fish stock often includes the analysis of fishing logs acquired through a logbook program. Data from these programs provide a history of relative catch and effort trends for a number of areas within the fishery and provide a record of marked change in those trends, should one occur. All vessels participating in the Shark Bay managed prawn fishery complete detailed logbooks providing information on the daily catch (kg) and amount of effort (hours trawled) expended in each of a number of fishing areas, which have been specified by the fishery scientists responsible for assessing the status of the fishery. Each boat provides the trawl duration and catch of every shot, i.e. haul of the trawl nets. King and tiger prawn stocks in the Shark Bay managed prawn fishery are monitored using the fishery-dependent catch and effort information from the log books (throughout the year), in conjunction with recruit surveys (March and April) and spawning stock surveys (June and July) for tiger prawns (Sporer et al., 2010). Recruit surveys are designed to assess the abundance of the young prawns that, through growth, will become available to the fishery later that year, whilst spawning stock surveys assess the abundance of mature female prawns that are able to reproduce. The recruitment surveys for tiger prawns also catch king prawns and thus contain information relevant to the stock status of the latter species. From these data, the catch per unit of effort (CPUE) can be determined for each fishing area. The king and tiger prawn stocks in the Shark Bay managed prawn fishery are believed to be fully exploited (Sporer & Kangas, 2004).

The standardised CPUE data from the fishery, also referred to as catch rates, are assumed to be an index of the abundance of the prawns that are vulnerable to the trawlers within that area. They are used as a means of tracking the performance of the fishery over successive seasons by monitoring changes in stock levels from year to year. Focusing on the main target species of the Shark Bay managed prawn fishery, this project aims to model the western king prawn catch rate distribution using a geostatistical spatiotemporal model, allowing for both spatial and temporal discrimination rather than the whole or part of region time series approach. The addition of spatial resolution to the analysis of CPUE in the prawn fishery would enable a better understanding of CPUE including a more realistic understanding of uncertainty, identification of areas of unprecedented or extended periods of localised depletion, and the ability to improve the timing of spatial openings and closures to maintain stocks or improve profitability.

Traditionally, geostatistics aims to use a sample data set of the spatial/spatiotemporal variable to model the spatial/spatiotemporal distribution of the variable and estimate or simulate

at unsampled spatial/spatiotemporal locations. Sample data for geostatistical analysis are often provided by surveys conducted for the purpose of the study. However, detailed survey data are not always available. This is particularly true for fisheries where obtaining fishery-independent survey data is costly. However, many fisheries have detailed fishery-dependent catch and effort data. As mentioned previously, the Shark Bay managed prawn fishery has detailed logbook catch and effort data provided by the various skippers of the fishing vessels in Shark Bay. In the absence of fishery-independent survey data throughout the season and across the entire spatial range of the Shark Bay managed prawn fishery, the fishery-dependent catch and effort data are utilised in this study.

Whilst logbook data are recorded in situ on the fishing vessel, time constraints of the fishing day and personal preferences may result in differences in the way the data are recorded by skippers on the various vessels. The prawn logbook data involve catch and effort data which, if considering a single fisherman, may typically be regarded as a non-random sample of the underlying prawn distribution as a fisherman targets known or preferred fishing areas. As such, the behaviour of the sample data of an individual fisherman may not completely represent, or may be a biased representation of, the behaviour over the study region. However, by considering the catch and effort of multiple fishermen in the same study, there is an element of randomness as their beliefs quite clearly do not coincide, otherwise they would all be fishing at the same location at a given time. This assumption validates the use of the random function model for the prawn catch rate data.

A pilot study involving king and tiger prawn catch data for the 2000 Shark Bay prawn fishing season showed the spatial aspect of the data was amenable to geostatistical methods (Mueller et al., 2004). A subsequent Fisheries Research and Development Corporation study (FRDC Project No. 2005/038) investigated the use of geostatistics to model the spatial distribution of Shark Bay prawn catch data (Mueller et al., 2008). The results of the study indicate that geostatistical techniques are suitable for analysing the prawn catch data. The FRDC project implemented the preliminary spatiotemporal modelling technique developed in this thesis to produce smoothed maps of king and tiger prawn catch rate distribution for a lunar month. The pictorial representations improved the understanding of the short-term and annual spatial variation in king prawn catch rate distribution and were seen as a useful tool in discussions with stakeholders on closure options (Mueller et al., 2008). The geostatistical spatiotemporal model for the prawn catch data is further developed in this thesis, including the

investigation of data aggregation by vessel and grid, the use of fishery-independent survey data to supplement the catch rate input data for estimation in a region as it is opened to fishing, and the addition of spatiotemporal simulation to capture the local uncertainty of the estimates.

The prawn catch data set is densely sampled in both the spatial and temporal dimensions and locations at successive time instants are not fixed. A suitable geostatistical spatiotemporal model must account for these properties. A spatially correlated time series approach is not directly applicable to these data due to the varying locations. Whilst the approach of temporally correlated (spatial) random functions is potentially valid, the vast number of time instants and varying spatial locations at each time instant is problematic. The single spatiotemporal random function model appears valid for analysing the prawn data by extracting the trend component from the prawn data and to analyse the resulting (stationary) spatiotemporal residuals. The presence and nature of temporal trends of the prawn catch data have been identified previously (Harman, 2001). Building on the previous spatial geostatistical analysis of king prawn catch rates, this study will assess the ability of a spatiotemporal model, by including the temporal dimension, to predict catch rates into the future, and provide a measure of the uncertainty associated with the estimates. An additional challenge faced with the prawn data is the spatiotemporal opening and closure lines implemented as part of the management regime in the fishery. Forecasting into a region just opened to fishing with no input data from previous weeks is problematic, as the catch rates peak in these new fishing regions. A region is typically targeted by fishers as it opens to take advantage of the increased number of prawns available. Catch rates from recruitment surveys are exploited in this study to condition the prediction of catch rates in these areas.

A geostatistical spatiotemporal model, based on the single spatiotemporal random function model, is implemented in this study with the king prawn catch and effort data from the Shark Bay managed prawn fishery to facilitate estimation and simulation of catch rates into the future. This model is data-driven and suited to the characteristics of the prawn catch rate data. The spatiotemporal analysis will provide insight into the spatiotemporal modelling ability of geostatistical methods to fishery-dependent catch and effort data allowing for the integration of limited fishery-independent survey data. This aims to reduce some of the knowledge gap that currently exists with regards to detailed spatiotemporal information about the king prawn distribution of Shark Bay. In doing so, this modelling technique may provide fishery managers with another tool for making better informed decisions.

1.2 OBJECTIVES

There are three aims of this study:

1. To develop a geostatistical spatiotemporal modelling framework for the Shark Bay western king prawn catch rate data based on the single spatiotemporal random function model.

   This approach will model and allow future predictions of the catch rate along with estimates of uncertainty.

2. To apply the developed geostatistical space-time model to the Shark Bay western king prawn catch rate data.

   Data from the 2001 to 2003 seasons will be used to compute estimates and simulated realisations for the 2004 season and then compared with the actual 2004 catch rate values.

3. To integrate fishery-independent survey data to improve the performance of estimates obtained for the Shark Bay western king prawn catch data from the developed geostatistical space-time model.

   Recruitment survey data will be used as supplementary information to improve catch rate estimates in a region as it is first opened to fishing in the 2004 season.

1.3 THESIS OUTLINE

This thesis is organised in four parts. In the first part, after a general introduction (Chapter 1), spatiotemporal geostatistical background theory is presented along with the methods of global trend modelling used in this study (Chapter 2). In Chapter 3 the king prawn catch rate data sets from the Shark Bay managed prawn fishery, which are used in the subsequent analysis, are presented. Then the methodology for space-time modelling of the king prawn data set is presented in Chapter 4, outlining the various components of the modelling process.

The second part of this thesis presents the global trend modelling process applied to the king prawn catch rate data (Chapter 5). This involves modelling the global multiplicative temporal trend of the averaged weekly catch rate by time series decomposition, involving a trend and cyclical component. The catch rate data are then transformed by dividing by the large-scale multiplicative temporal trend. The spatial correlation and continuity over time of the standardised king prawn catch rate data are presented in Chapter 6. In Chapter 7, the

spatiotemporal continuity of the standardised king prawn catch rate data is analysed and modelled via variography for use in the subsequent estimation and simulation methods.

The third part of this thesis presents the spatiotemporal geostatistical estimation (Chapter 8) and simulation (Chapter 9) results of the 2004 king prawn catch rate data. These analyses are conducted using the 2001 to 2003 data as input data. The standardised catch rate data are estimated via a space-time kriging procedure and are then back transformed to catch rate values by incorporating the large-scale multiplicative temporal trend. Estimates in a region as its opened to fishing are improved using survey data as conditioning data in addition to the catch rate input data. Simulation is performed to characterise the local uncertainty of the catch rate estimates.

The final part of this thesis (Chapter 10) presents a summary of this study along with possible research avenues for the continued, and improved, spatiotemporal geostatistical modelling of the king prawn distribution, along with the potential modelling of the distributions of other prawn species.

CHAPTER 2

THEORETICAL FRAMEWORK

2.1 INTRODUCTION

The main focus of this research is the development of a spatiotemporal geostatistical modelling approach for the Shark Bay king prawn catch rate data. This chapter outlines the geostatistical modelling framework for spatiotemporal data, including the data framework, modelling of spatiotemporal continuity along with possible estimation and simulation methods. Geostatistical models of spatiotemporal variability assume stationarity throughout the domain, so that variables require global detrending before analysis. The methods of global trend modelling used in this study (time series and median polish) are also presented in this chapter.

2.2 SPATIOTEMPORAL GEOSTATISTICS

Geostatistics utilises the theory of random functions to model regionalised variables. A variable is said to be regionalised when it is distributed in space and/or time. The regionalised variable (abbreviated to ReV) is usually a characteristic of a certain phenomenon which we call the regionalisation. In most cases, a characteristic behaviour or structure of the spatiotemporal variability of the ReV under study can be detected behind a locally erratic aspect. Thus, a ReV possesses two apparently contradictory characteristics,

1. A local, random, erratic aspect.
2. A general structured aspect which requires a functional representation.

Geostatistics is a set of techniques that use a probabilistic interpretation to take into account both aspects of randomness and structure of ReVs. The mathematical construct that is used to model the ReV is a random function (abbreviated to RF). In geostatistics we act as though the ReV under study is a realisation of a parent RF. Geostatistics can thus be defined as the application of the probabilistic theory of RFs to ReVs. Traditionally geostatistics has been concerned with spatially varying variables, but has been extended to involve variables varying in both space and time (Kyriakidis & Journel, 1999a). The philosophy is, where data are collected in positions and at varying times within a region, these locations and time instants may help to explain their variability. We will refer to the traditional methods of geostatistics as spatial geostatistics and use the more general term geostatistics to refer to spatiotemporal geostatistics, which is the focus of this thesis.

2.2.1 THE JOINT SPACE-TIME FRAMEWORK

In order to analyse observations of a ReV, varying in both space and time, we place them into a spatiotemporal framework generalised from the spatial framework of Goovaerts (1997). The combined space-time domain is denoted $\mathcal{A} \subseteq D \times T$ where D is a finite domain in space with $D \subseteq \mathbb{R}^d$ and T is a finite domain in time with $T \subseteq \mathbb{R}^+$, and often $T \subseteq \mathbb{Z}^+$. The coordinates of a spatiotemporal location are denoted by $(\mathbf{u},t)$ where $\mathbf{u} \in D$ and $t \in T$, and the value of the attribute at that spatiotemporal location is denoted by $z(\mathbf{u},t)$. Note that the spatial locations are continuous whereas the time instants may be continuous or discrete, and therefore the number of spatiotemporal locations is infinite. The number of these spatiotemporal locations at which the attribute value is known is finite and denoted by n_α. These are called sampled spatiotemporal locations.

2.2.2 SINGLE SPATIOTEMPORAL RANDOM FUNCTION MODEL

We can consider the data value $z(\mathbf{u},t)$ as a realisation of a random variable $Z(\mathbf{u},t)$ at the spatiotemporal location $(\mathbf{u},t)$. A random variable (abbreviated to RV) is discrete if it has a range with only a finite or countably infinite number of values or continuous if it has a range with an uncountably infinite number or values on the real line. As we will be concerned with continuous RVs in our study let us use the term RV to describe a continuous RV. The behaviour of a RV is characterised by its probability distribution.

Definition 2.1: Cumulative Distribution Function

The cumulative distribution function (cdf) of a RV is given by

$$F(\mathbf{u},t;z) = \Pr\{Z(\mathbf{u},t) < z\} \tag{2.1}$$

for all z.

The cdf is the probability that the value of $Z(\mathbf{u},t)$ is less than or equal to a specific value *z* and models the uncertainty of the value $z(\mathbf{u},t)$. We can assign a RV $Z(\mathbf{u},t)$ to each location in the study area *A*. The set of all these RVs is known as a random function (abbreviated to RF). For short the RF is also denoted by $Z(\mathbf{u},t)$ like the RV but it is typically easy to distinguish which one is being used.

In geostatistics it is assumed that the ReV $\{z(\mathbf{u},t);(\mathbf{u},t) \in \mathcal{A}\}$ under study is a realisation of a parent RF $\{Z(\mathbf{u},t);(\mathbf{u},t) \in \mathcal{A}\}$, which is a collection of realisations of its components RVs. The RF is denoted by $\mathbf{Z}(\mathbf{u},t)$. For every set of *k* points in $\mathcal{A}$, there corresponds a *k*-component

vectorial RV $\{Z((\mathbf{u},t)_1), Z((\mathbf{u},t)_2),..., Z((\mathbf{u},t)_k)\}$. This vectorial RV is characterised by the *k*-variable joint distribution function.

Definition 2.2: Joint Cumulative Distribution Function

The joint cdf of a RF is given by

$$F[(\mathbf{u},t)_1,(\mathbf{u},t)_2,...,(\mathbf{u},t)_k; z_1, z_2,..., z_k] = \Pr\{Z(\mathbf{u},t)_1 < z_1, Z(\mathbf{u},t)_2 < z_2,...,Z(\mathbf{u},t)_k < z_k\} \quad (2.2)$$

where $(\mathbf{u},t)_k \in \mathcal{A}$ *for all possible choices of k. This joint cdf characterises the joint uncertainty about the k actual values. The set of all these distribution functions, for all positive integers k and for every possible choice of support points in* $\mathcal{A}$*, constitutes the spatiotemporal law of the RF.*

In practice, we limit ourselves to cdfs involving no more than two locations at a time which are outlined as follows. The ***univariate (one-point) cdf*** is given by

$$F(\mathbf{u},t;z) = \Pr\{Z(\mathbf{u},t) \le z\} = E\{I(\mathbf{u},t;z)\} \quad (2.3)$$

and the bivariate (two-point) cdf is given by

$$\begin{aligned} F[(\mathbf{u},t),(\mathbf{u},t)';z,z'] &= \Pr\{Z(\mathbf{u},t) \le z, Z((\mathbf{u},t)') \le z'\} \\ &= E\{I(\mathbf{u},t;z) \cdot I((\mathbf{u},t)';z')\} \end{aligned} \quad (2.4)$$

where for all *z* and all $(\mathbf{u},t)$, $I(\mathbf{u},t;z)$ denotes the binary random variable

$$I(\mathbf{u},t;z) = \begin{cases} 1, & \text{if } Z(\mathbf{u},t) \le z \\ 0, & \text{otherwise} \end{cases} \quad (2.5)$$

The one-point and two-point terminology indicates that the two random variables relate to the same attribute *z* at two different locations and not to two different variables. To infer behaviour of the RV $Z(\mathbf{u},t)$ we also consider the first and second order moments of the RV.

Definition 2.3: Moment of a Random Function

The joint cdf of a RF,given by the kth order moment of the RV $Z(\mathbf{u},t)$ *at any location* $(\mathbf{u},t)$*, is defined as*

$$E[Z(\mathbf{u},t)]^k = \int x^k dF(\mathbf{u},t;x) \quad (2.6)$$

provided this integral exists where E[.] denotes the expected value of the random variable.

The ***expected value*** of a random variable is defined to be its first order moment:

$$m(\mathbf{u},t) = E\left[Z(\mathbf{u},t)\right] \tag{2.7}$$

The expected value in general is allowed to depend on the spatiotemporal location $(\mathbf{u},t)$. In geostatistics we often refer to the expectation as the trend which represents the large-scale changes of $Z(\mathbf{u},t)$. The ***variance*** of a RF $Z(\mathbf{u},t)$ is defined as the second-order moment about the expectation $m(\mathbf{u},t)$:

$$Var\left[Z(\mathbf{u},t)\right] = E\left[Z(\mathbf{u},t) - m(\mathbf{u},t)\right]^2 \tag{2.8}$$

for any location $(\mathbf{u},t)$. As for the trend, the variance is generally dependent on the location $(\mathbf{u},t)$.

An important variant of the second-order moment, the (two-point) ***covariance*** is defined as

$$C\left[(\mathbf{u},t),(\mathbf{u},t)'\right] = E\left[Z(\mathbf{u},t)\cdot Z\left((\mathbf{u},t)'\right)\right] - E\left[Z(\mathbf{u},t)\right]\cdot E\left[Z\left((\mathbf{u},t)'\right)\right] \tag{2.9}$$

for any locations $(\mathbf{u},t)$ and $(\mathbf{u},t)'$. When $(\mathbf{u},t) = (\mathbf{u},t)'$ the covariance equals the variance of $(\mathbf{u},t)$. The covariance structure of the RV $Z(\mathbf{u},t)$ represents its variability due to small and microscale stochastic sources. The (two-point) ***correlogram*** is the standardised form of covariance given by

$$\rho\left[(\mathbf{u},t),(\mathbf{u},t)'\right] = \frac{C\left[(\mathbf{u},t),(\mathbf{u},t)'\right]}{\sqrt{C\left[(\mathbf{u},t),(\mathbf{u},t)\right]\cdot C\left[(\mathbf{u},t)',(\mathbf{u},t)'\right]}} \tag{2.10}$$

The (semi)variogram is an alternative way of describing the second order properties of a RV. In traditional statistics, the dependency between RVs is expressed in terns of covariances whilst in geostatistics it is more common to work with (semi)variograms. The ***variogram*** is the variance of the difference of the two RVs defined by the locations $(\mathbf{u},t)$ and $(\mathbf{u},t)'$ given by

$$2\gamma\left[(\mathbf{u},t),(\mathbf{u},t)'\right] = Var\left[Z(\mathbf{u},t) - Z\left((\mathbf{u},t)'\right)\right] \tag{2.11}$$

The variogram describes how this variance changes with changes in the separation distance between the two locations.

The semivariogram $\gamma\left[(\mathbf{u},t),(\mathbf{u},t)'\right]$, is perhaps the simplest way to relate uncertainty with distance and/or time from an observation and is the most traditional and commonly used inferential tool of spatial/spatiotemporal geostatistics for specifying the spatial/spatiotemporal

structure. The semivariogram estimator calculates the dissimilarity of a data pair as half the squared difference between the observed values. The average of the pairwise dissimilarities over a set of lags results in the experimental semivariogram. For a stationary process, the mean and variance of the observed phenomenon is identical for any subregion of the domain, and the semivariogram can be used as a valid description of the spatiotemporal continuity of the variable. This means that the difference between any pair of data points at arbitrary (spatiotemporal) separation is purely a function of their (spatiotemporal) separation vector.

A measure of dependence between $Z(\mathbf{u},t)$ and $Z(\mathbf{u},t)'$ is the centred two-point cdf, defined as

$$F\left[(\mathbf{u},t),(\mathbf{u},t)';z,z'\right]-F\left[(\mathbf{u},t);z\right]\cdot F\left[(\mathbf{u},t)';z'\right] \tag{2.12}$$

The two RVs are independent if expression (2.12) equals zero. Using (2.5) we can rewrite this dependence relation as

$$E\left[I(\mathbf{u},t;z)\cdot I\left((\mathbf{u},t)';z'\right)\right]-E\left[I(\mathbf{u},t;z)\right]\cdot E\left[I\left((\mathbf{u},t)';z'\right)\right]=C_I\left[(\mathbf{u},t),(\mathbf{u},t)';z,z'\right] \tag{2.13}$$

The set of all indicator cross covariances $C_I\left[(\mathbf{u},t),(\mathbf{u},t)';z,z'\right]$ for all thresholds provides a measure of dependence between the two RVs $Z(\mathbf{u},t)$ and $Z(\mathbf{u},t)'$. The Z-covariance (2.9) is a measure only of linear correlation between the two RVs. The Z-covariance and all indicator cross covariances vanish when the two RVs are independent. However the condition $C\left[(\mathbf{u},t),(\mathbf{u},t)'\right]=0$ does not necessarily imply (2.12) equals zero. Two linearly uncorrelated RVs may still be dependent, in which case the dependence relation is non-linear (Goovaerts, 1997).

2.2.3 STATIONARITY

Inference of the spatiotemporal law requires repeated realisations of the component RVs at each space-time location $(\mathbf{u},t)\in\mathcal{A}$, which are never available in practice. Even if a very large of number of realisations were available, the combinatorial possibilities are such that we could only calculate multidimensional distributions only for the simplest *k*-tuples. In our instance we only have a single (partial) realisation of the RF and therefore cannot infer the entire spatiotemporal law of (2.2) (Kyriakidis & Journel, 1999a). Even the more practical inference of the cdf and moment relations of (2.3) - (2.11) requires multiple realisations at each location **u**. For example, inference of the Z-covariance $C\left[(\mathbf{u},t),(\mathbf{u},t)'\right]$ defined in (2.9) between the two RVs $Z(\mathbf{u},t)$ and $Z\left((\mathbf{u},t)'\right)$separated by the vector $\mathbf{h}=(\mathbf{h}_s,h_t)=(\mathbf{u},t)-(\mathbf{u},t)'$ requires a set of repeated measurements

$$\left\{z^{(l)}(\mathbf{u},t), z^{(l)}((\mathbf{u},t)'); l=1,\ldots,L\right\} \tag{2.14}$$

that are never available in practice. Simplification is needed and it is provided by the assumption of stationarity. Stationarity allows us to replace the unavailable repetition with repetition associated with pairs of locations. It must be noted that stationarity is a property of the RF model. It is not a characteristic of the phenomenon under study but rather a decision made by the user.

Under the assumption of stationarity, pairs of values separated by the space-time vector $\mathbf{h}$ within the study area $\mathcal{A}$ are used as a set of repeated measurements

$$\left\{z(\mathbf{u},t)_\alpha, z((\mathbf{u},t)_\alpha + \mathbf{h}); \alpha=1,\ldots,n\right\} \tag{2.15}$$

The implicit assumption is that the corresponding pairs of RVs

$$\left\{Z(\mathbf{u},t)_\alpha, Z((\mathbf{u},t)_\alpha + \mathbf{h}); \alpha=1,\ldots,n\right\} \tag{2.16}$$

originate from the same two-point distribution. This provides the many realisations of the RF $\mathbf{Z}(\mathbf{u},t)$ necessary to make statistical inference. Such pooling of data pairs regardless of their locations calls for the phenomenon under study to be homogeneous within the study area. In probabilistic terms, the RF model must be chosen to be stationary within $\mathcal{A}$. Strict stationarity requires the spatiotemporal law of the spatiotemporal variable $\mathbf{Z}(\mathbf{u},t)$ to be invariant under translation.

In practice, we only need to justify that the variable is stationary up to a given distance and within a given time frame. If a variable is deemed locally stationary within a spatial region such that the fluctuations of the variable dominate any trend then we can justify the use of spatial geostatistical tools within this neighbourhood (Armstrong, 1998). We can extend this principle to apply spatiotemporal geostatistical tools to a spatiotemporal variable deemed stationary within a spatial region and a particular time frame.

Definition 2.4: Strictly Stationary Random Function

The RF $\mathbf{Z}(\mathbf{u},t)$ *is said to be strictly stationary within* $\mathcal{A}$ *if the multivariate cdf is invariant under translation, given by*

$$\begin{aligned} F\left[(\mathbf{u},t)_1,(\mathbf{u},t)_2,\ldots,(\mathbf{u},t)_N; z_1,z_2,\ldots,z\right]_N = F\left[(\mathbf{u},t)_1+\mathbf{h},(\mathbf{u},t)_2+\mathbf{h},\ldots,(\mathbf{u},t)_N+\mathbf{h};\right. \\ \left. z_1,z_2,\ldots,z_N\right] \quad \forall (\mathbf{u},t)_1,\ldots,(\mathbf{u},t)_N \text{ and } \mathbf{h} \end{aligned} \tag{2.17}$$

Invariance of the multivariate cdf entails invariance of any lower order cdf, including the univariate and bivariate cdfs, and invariance of all their moments including covariances. This property is usually difficult to test and cannot be verified from the limited sample data as it needs to be demonstrated by considering the family of finite-dimensional distribution functions of the process (Bruno & Cocchi, 2004). Furthermore, it is often an unreasonable assumption for the data at hand. For this reason, weaker forms of stationarity may be sufficient to provide a foundation for modelling analysis. The assumption of strict stationarity can be loosened by limiting the decision of stationarity to the one-point and two-point cdfs and the first two moments of the RF (Goovaerts, 1999; Kyriakidis, 1999).

This less strict definition of stationarity is known as second-order stationarity. For a second-order stationary RF, the reference to a particular spatiotemporal location $(\mathbf{u},t)$ can be dropped from the expressions of (2.3) - (2.11). The one-point and two-point cdfs and corresponding moments now depend only on the separation vector $\mathbf{h}=(\mathbf{h}_s,h_t)$ between the locations $(\mathbf{u},t)$ and $(\mathbf{u},t)'$:

- ***Univariate (one-point) cdf***

$$F(z)=F(\mathbf{u},t;z)=\Pr\{Z(\mathbf{u},t)\le z\} \tag{2.18}$$

- ***Bivariate (two-point) cdf***

$$F(\mathbf{h};z,z')=\Pr\{Z(\mathbf{u},t)\le z, Z(\mathbf{u}+\mathbf{h}_s,t+h_t)\le z'\} \tag{2.19}$$

- ***Z-expected value***

$$m=E[Z(\mathbf{u},t)] \tag{2.20}$$

- ***Z-covariance (two-point)***

$$C(\mathbf{h})=E[Z(\mathbf{u},t)\cdot Z(\mathbf{u}+\mathbf{h}_s,t+h_t)]-E[Z(\mathbf{u},t)]\cdot E[Z(\mathbf{u}+\mathbf{h}_s,t+h_t)] \tag{2.21}$$

- ***Z-correlogram (two-point)***

$$\rho(\mathbf{h})=\frac{C(\mathbf{h})}{C(\mathbf{0})} \tag{2.22}$$

- ***Z-variogram***

$$2\gamma(\mathbf{h})=Var[Z(\mathbf{u})-Z(\mathbf{u}+\mathbf{h}_s,t+h_t)]=E\{[Z(\mathbf{u})-Z(\mathbf{u}+\mathbf{h}_s,t+h_t)]^2\} \tag{2.23}$$

The separation vector $\mathbf{h} = (\mathbf{h}_s, h_t)$ accounts for both distance $|\mathbf{h}|$ and direction. $C(\mathbf{h})$ and $2\gamma(\mathbf{h})$ are said to be anisotropic if they depend on both distance and direction. They are isotropic if they depend only on the modulus of **h**. A second-order stationary RF is defined by its expectation and covariance:

Definition 2.5: Second Order Stationary Random Function

The RF $\mathbf{Z}(\mathbf{u})$ *is said to be second-order stationary within* $\mathcal{A}$ *if the first moment of the vector of random variables exists and is constant and the covariance of all pairs of RVs exists and depends only on the separation vector*

$$E\{Z(\mathbf{u},t)\} = m \tag{2.24}$$

$$E\{[Z(\mathbf{u},t) - m][Z(\mathbf{u}+\mathbf{h}_s, t+h_t) - m]\} = C(\mathbf{h}) \tag{2.25}$$

Strict stationarity implies second-order stationarity, assuming that the first two moments exist, whereas second-order stationarity does not imply strict stationarity. In a Gaussian spatiotemporal process, the second order stationarity and strict stationarity coincide, as a Gaussian process is completely specified by its mean and variance (Bruno & Cocchi, 2004).

In many cases the assumption of second-order stationarity may not be met. A weaker assumption of intrinsic stationarity assumes that the increments $Z(\mathbf{u},t) - Z(\mathbf{u}+\mathbf{h}_s, t+h_t)$ themselves are second-order stationary (Journel & Huigbregts, 1978; Goovaerts, 1997). Intrinsic stationarity is a sufficient condition to define the semivariogram.

Definition 2.6: Intrinsically Stationary Random Function

The RF $Z(\mathbf{u},t)$ *is said to be intrinsically stationary within* $\mathcal{A}$ *if the mean and variance of the increments* $Z(\mathbf{u},t) - Z(\mathbf{u}+\mathbf{h}_s, t+h_t)$ *exist and are translation invariant.*

$$E\{Z(\mathbf{u},t) - Z(\mathbf{u}+\mathbf{h}_s, t+h_t)\} = 0 \tag{2.26}$$

$$Var\{Z(\mathbf{u},t) - Z(\mathbf{u}+\mathbf{h}_s, t+h_t)\} = E\left\{\left[Z(\mathbf{u},t) - Z(\mathbf{u}+\mathbf{h}_s, t+h_t)\right]^2\right\} = 2\gamma(\mathbf{h}) \tag{2.27}$$

There is a relationship between the covariance, correlogram and SV of a stationary RF.

$$\gamma(\mathbf{h}_s, h_t) = C(\mathbf{0},0) - C(\mathbf{h}_s, h_t) \tag{2.28}$$

$$\rho(\mathbf{h}_s, h_t) = 1 - \frac{\gamma(\mathbf{h}_s, h_t)}{C(\mathbf{0},0)} \tag{2.29}$$

As the separation distance $|\mathbf{h}_s|$ or separation time $|h_t|$ increases, the correlation between any two RVs $Z(\mathbf{u},t)$ and $Z(\mathbf{u}+\mathbf{h}_s, t+h_t)$ generally tends to zero.

$$C(\mathbf{h}_s, h_t) \to 0 \quad \text{for} \quad |\mathbf{h}_s| \to \infty \quad \text{or} \quad |h_t| \to \infty \tag{2.30}$$

The sill value of a bounded SV tends toward the a priori variance $C(\mathbf{0})$.

$$\gamma(\mathbf{h}_s, h_t) \to C(\mathbf{0},0) \quad \text{for} \quad |\mathbf{h}_s| \to \infty \quad \text{or} \quad |h_t| \to \infty \tag{2.31}$$

2.2.4 NON-STATIONARITY

The stationarity decision allows pooling data over areas that are deemed homogeneous. In practice, many data sets do not satisfy the assumption of second order or intrinsic stationarity. Furthermore, the departure from stationarity is often linked to the presence of a trend and therefore it is necessary to remove the trend from the observation via some model. These trends can either be removed from the data and added back to the interpolated estimates of the residuals, or incorporated directly into the interpolation scheme. Once the global deterministic trend is removed, uncertainties in the detrending procedure are not taken into account in further analysis (Snepvangers et al., 2003).

The possible sources of non-stationarity that exist within a data set are a non-constant mean, a non-constant variance or a spatiotemporally varying covariance function (Elmatzoglou, 2006). All sources of non-stationarity must be resolved in order to model the spatiotemporal variability via a (second-order) stationary spatiotemporal random function model. A non-constant mean can be accommodated by expressing the mean function in terms of the spatiotemporal coordinates or other related variables. A non-constant variance can be stabilised by transformation of the response variable. A spatially and/or temporally varying covariance function cannot easily be dealt with, and is not examined in this study.

In some cases, even if data are non-stationary, they may be considered locally stationary (that is, to exhibit no definable trend) over a shorter distance, called the limit of quasi-stationarity (Armstrong, 1998). Subsequent modelling (and estimation or simulation) can be performed over this limit, rather than relying on a trend function. This is achievable in all kriging methods using an appropriate search neighbourhood.

Exploratory data analysis may suggest the existence of several populations with significantly different statistics. In this case, the region should be subdivided into more homogeneous zones with each being modelled with a different RF (Goovaerts, 1999). This

subdivision is dependent on the availability of enough data to infer the parameters of each separate RF and the ability to delineate the different populations for both data and unsampled locations.

2.2.5 SEMIVARIOGRAM AND COVARIANCE INFERENCE

Once a random function model has been chosen, the next step is to infer its parameters from the available information (sample data). The inference process aims at estimating the parameters of the RF model from the sample information available over the study area. Inference of the first two moments (mean, covariance) of the RF $Z(\mathbf{u},t)$ is typically the focus of geostatistical modelling as they are required by subsequent estimation or simulation algorithms. For convenience, we often work with semivariograms rather than covariances. Theoretical models of the spatiotemporal continuity of an attribute are required in order to derive values of the semivariogram for any separation vector **h**.

In traditional (spatial) geostatistics, the experimental spatial semivariogram $\hat{\gamma}_s(\mathbf{h}_s)$ is the primary tool of inference. Similarly, the experimental spatiotemporal semivariogram $\hat{\gamma}_{st}(\mathbf{h}_s,h_t)$ is the primary tool for inference in spatiotemporal geostatistics and is defined as

$$\hat{\gamma}_{st}(\mathbf{h}_s,h_t)=\frac{1}{2N(\mathbf{h}_s,h_t)}\sum_{\alpha=1}^{N(\mathbf{h}_s,h_t)}\left[z(\mathbf{u},t)_\alpha-z\left((\mathbf{u},t)_\alpha+(\mathbf{h}_s,h_t)\right)\right]^2 \tag{2.32}$$

where $N(\mathbf{h}_s,h_t)$ is the number of pairs of sample locations separated by the vector $\mathbf{h}=(\mathbf{h}_s,h_t)$, $\mathbf{h}_s$ is the space lag, h_t is the time lag and $z(\mathbf{u},t)_\alpha$ is the value of the variable at the spatiotemporal location $(\mathbf{u},t)_\alpha$.

The experimental spatiotemporal covariance $\hat{C}_{st}(\mathbf{h})$ is defined as

$$\hat{C}_{st}(\mathbf{h}_s,h_t)=\frac{1}{2N(\mathbf{h}_s,h_t)}\sum_{\alpha=1}^{N(\mathbf{h}_s,h_t)}z_i\left((\mathbf{u},t)_\alpha\right)\cdot z\left((\mathbf{u},t)_\alpha+(\mathbf{h}_s,h_t)\right)-\hat{m}_{-h}\cdot\hat{m}_{+h} \tag{2.33}$$

with

$$\hat{m}_{-\mathbf{h}}=\frac{1}{2N(\mathbf{h})}\sum_{\alpha=1}^{N(\mathbf{h})}z(\mathbf{u}_\alpha),\qquad \hat{m}_{+\mathbf{h}}=\frac{1}{2N(\mathbf{h})}\sum_{\alpha=1}^{N(\mathbf{h})}z(\mathbf{u}_\alpha+\mathbf{h}) \tag{2.34}$$

where $\hat{m}_{-\mathbf{h}}$ is the mean of the tail values and $\hat{m}_{+\mathbf{h}}$ is the mean of the head values.

For a finite number of separation distances $\mathbf{h}_j, j=1,\ldots,J$ (in various directions if necessary), a set of experimental values $\hat{\gamma}(\mathbf{h}_j)$ is obtained from the sample data and these values are plotted against $|\mathbf{h}|$. Sample variograms do not provide all of the separation distances

and the corresponding semivariances needed by the kriging system. It is necessary to have a model that enables computing a variogram value for any possible separation distance, so continuous functions must be fitted to the experimental values. A theoretical model that best captures the overall features of the experimental semivariogram is then chosen. Only certain functions, or combination of functions, may be used to model the theoretical covariance or semivariogram.

2.2.6 PERMISSIBLE MODELS

A finite linear combination *Y* of *n* random variables is a random variable whose variance must be nonnegative. The variance of this linear combination may be expressed as a linear combination of covariances

$$Var(Y) = Var\left(\sum_{\alpha=1}^{n} \lambda_\alpha Z\left[(\mathbf{u},t)_\alpha\right] \right) = \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta C\left[(\mathbf{u},t)_\alpha - (\mathbf{u},t)_\beta\right] \geq 0 \tag{2.35}$$

for any set of *n* locations $(\mathbf{u},t)_\alpha$ and any weights λ_α. To ensure that this variance is non-negative, the covariance model $C(\mathbf{h}_s, h_t)$ must be positive definite. Accounting for the relation (2.28), the variance (2.35) is rewritten in terms of the semivariogram model $\gamma(\mathbf{h},t)$. The semivariogram model must then be conditionally negative definite to ensure the non-negativity of the variance, with the condition being that the sum of the weights λ_α is zero.

$$Var(Y) = C(\mathbf{0},0)\sum_{\alpha=1}^{n}\lambda_\alpha \sum_{\beta=1}^{n}\lambda_\beta - \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n} \lambda_\alpha \lambda_\beta \gamma\left[(\mathbf{u},t)_\alpha - (\mathbf{u},t)_\beta\right] \geq 0 \tag{2.36}$$

2.2.7 SPATIAL AND TEMPORAL COVARIANCE/SEMIVARIOGRAM MODELS

Many of the spatiotemporal covariance/semivariogram models require the specification of the associated marginal spatial and temporal covariance/semivariogram models or use permissible spatial covariance/semivariogram models incorporating an additional dimension for time. As the conditions of positive/negative definiteness are not easy to check, in practice, we choose from a number of parametric models that are known to satisfy these principles. There are a number of basic models that are known to be permissible for either the spatial and/or temporal domain (Goovaerts, 1997; Deutsch & Journel, 1998). A few of the possible models are expressed here in their isotropic form as a function of the scalar $h = |\mathbf{h}|$ where $\mathbf{h}$ is the spatial or temporal separation vector and with graphical comparison. These are the models employed in this study, so this list is by no means exhaustive.

- ***Nugget Effect model***

$$\gamma_{Nug}(h) = \begin{cases} 0, & \text{if } h = 0 \\ b, & \text{if } h > 0 \end{cases} \tag{2.37}$$

- ***Spherical Model with range a***

$$\gamma_{Sph}(h) = \begin{cases} b\left(1.5 \cdot \frac{h}{a} - 0.5 \cdot \left(\frac{h}{a}\right)^3\right), & \text{if } h \le a \\ b, & \text{if } h > a \end{cases} \tag{2.38}$$

- ***Exponential Model with practical range a***

$$\gamma_{Exp}(h) = b\left(1 - \exp\left(\frac{-3h}{a}\right)\right) \tag{2.39}$$

Features of the semivariogram model can be used to interpret the behaviour of the spatial continuity of the variable under study. The relative nugget is the ratio of the nugget to the total sill of the variogram and defines the proportion of randomness in the spatial continuity of the variable, represented by a discontinuity in the semivariogram at the origin (Isaaks & Srivastava, 1989). A variable with a weaker spatial structure will exhibit a higher relative nugget, whilst a lower relative nugget indicates a stronger spatial structure. The relative nugget tends to increase with the lag tolerance and with data sparsity. The relative nugget typically decreases with the use of more data or data of an increased quality. Where data are clustered and there is a proportional effect, the relative nugget effect is better inferred from relative semivariograms. (Goovaerts, 1997)

The distance where the model flattens out is known as the range. Locations separated by distances closer than the range are spatially autocorrelated, whereas locations farther apart than the range are not. The value that the semivariogram model attains at the range is called the sill. The partial sill is the sill minus the nugget.

2.2.7.1 ANISOTROPIC MODELS

A phenomenon is said to be anisotropic when its pattern of spatial continuity changes with direction. It makes no sense to speak of anisotropy in time as there is only one temporal dimension. Modelling anisotropy calls for functions that depend on the vector $\mathbf{h}$ rather than on the distance $h = |\mathbf{h}|$ only. There are two types of anisotropy; geometric anisotropy and zonal anisotropy. Geometric anisotropy occurs when the directional semivariograms (covariances)

have the same shape and sill but different range values. Zonal anisotropy involves sill values varying with direction.

2.2.7.2 LINEAR MODEL OF REGIONALISATION

In most situations, two or more basic models must be combined to fit the shape of the experimental semivariogram or covariance function. However not all combinations of permissible semivariogram or covariance models result in a permissible semivariogram or covariance model. Sufficient conditions for a linear covariance model to be a permissible model of regionalisation are that each of the basic functions is a permissible covariance model and that the sill of each basic covariance model is positive.

2.2.7.3 WEIGHTED LEAST-SQUARES CRITERIA

The semivariogram modelling process is not just an exercise in curve fitting as there is always uncertainty attached to the parameters of the semivariogram model. As many models can appear to match equally well the sample information, there is some argument over the best way to proceed (Goovaerts, 1997). The first one is manual fitting, in which a theoretical semivariogram model is selected based on visual inspection of the empirical semivariogram. The user will decide whether to fit an isotropic or anisotropic model, which number and type of basic semivariogram models, along with the parameters of each basic semivariogram model to use. The second approach is to perform model fitting in an automated manner using methods such as least squares, maximum likelihood, and robust methods (Cressie, 1993). In this study we favour manual fitting. However, statistical criteria can be used to help justify the choice of a particular model and its set of parameters. A commonly used criterion used is the weighted least squares given by

$$W(\psi) = \sum_{N(\mathbf{h})} \omega(\mathbf{h}) \cdot \left[\hat{\gamma}(\mathbf{h}) - \gamma(\mathbf{h};\psi)\right]^2 \tag{2.40}$$

or

$$W(\psi) = \sum_{N(\mathbf{h})} \omega(\mathbf{h}) \cdot \left[\frac{\hat{\gamma}(\mathbf{h})}{\gamma(\mathbf{h};\psi)} - 1\right]^2 \tag{2.41}$$

where $\omega(\mathbf{h})$ is the weight given to each lag $\mathbf{h}$, $\hat{\gamma}(\mathbf{h})$ is the semivariogram estimate calculated from sample data using (2.32) and $\gamma(\mathbf{h};\psi)$ is the semivariogram model value for the associated set of parameters ψ. The weight $\omega(\mathbf{h})$ is often taken proportional to the number of data pairs $N(\mathbf{h})$ that contribute to the estimate $\hat{\gamma}(\mathbf{h})$ (Goovaerts, 1997; De Iaco, 2010). Automatic fitting

procedures using least squares select the most appropriate model to minimise the weighted sum of squares. In this study we favour manual fitting due to the erratic semivariograms. However, the weighted sum of squares can be used to compare the fits of various models for a particular season and to compare the fits of various seasons. Statistical criteria help justify the choice of a particular model and its set of parameters.

2.2.8 SPATIOTEMPORAL COVARIANCE/SEMIVARIOGRAM MODELS

Estimating and modelling the correlation of a space-time process is fundamental to geostatistical analysis, since only if the correlation model is appropriate for the variable under study can the subsequent estimation and/or simulation results be relied on (De Iaco, 2010). The semivariogram can be modelled with any conditionally negative definite function. Whilst it is possible to propose a model and test its permissibility, in practice it is customary to use one of a set of basic models that are known to be permissible. This is a relatively straightforward practice in modelling spatial (or temporal) variograms as there are many such models in common use and these models may be combined linearly to form complex models. However, the use of spatiotemporal covariance/semivariogram models is less common, although it has become an active area or research in the past decade or so. Different authors have suggested a large number of spatiotemporal covariance/semivariogram models (Dimitrakopoulos & Luo, 1994; Cressie & Huang, 1999; Kyriakidis & Journel, 1999a; De Iaco et al., 2001; De Iaco et al., 2002a; Gneiting, 2002; Ma, 2003, 2005; Stein, 2005). A comparative review of the characteristics, benefits and shortcomings of many of these currently accepted and implemented models has recently been undertaken (De Iaco, 2010).

One of the main distinctions made between spatiotemporal covariance/semivariogram models is based on the notion of separability. A *separable* space–time covariance function considers the spatiotemporal process as the joint process of two independent processes, one that occurs in space and another that occurs in time, resulting in a purely spatial component and a purely temporal component. However, we do not observe realisations of the two separate processes, only the joint process. This formulation allows for computationally efficient estimation and inference, which has led to separable covariance models being used even in situations in which they are not physically justifiable (Cressie & Huang, 1999; Bruno, 2004). Separability is restrictive and often requires unrealistic assumptions. Studies have suggested ways of testing for separability (Mitchell et al., 2005; Fuentes, 2006; Bevilacqua et al., 2010).

2.2.8.1 METRIC MODEL

The metric space-time covariance model (Dimitrakopoulos & Luo, 1994; Soares & Pereira, 2007) is given by

$$C_{st}(\mathbf{h}_s, h_t) = C\left(a_1 \left|\mathbf{h}_s\right|^2 + a_2 \left|h_t\right|^2\right) \tag{2.42}$$

where the coefficients $a_1, a_2 \in \mathbb{R}^+$ enable the comparison between space and time. Introducing a norm or metric in space-time to compare distance in space with distance in time has been criticised because it ignores the fundamental difference between space and time (Myers, 2002). The use of parameters a_1 and a_2 is analogous to the use of geometric anisotropy in spatial geostatistics, where time is considered as another dimension to the 2 or 3 spatial dimensions. Variograms are computed for the spatial dimension by considering spatial lag pairs for zero temporal separation. Similarly, the temporal variogram is computed for zero spatial separation, which is possible for fixed spatial locations. If locations vary over time, then the temporal variogram may need to be computed for a pseudo zero lag spacing using a very small separation tolerance.

In the metric model, the same type of model is assumed for the spatial and temporal covariances, with possible changes in the range (Figure 2.1). A pure metric model is restrictive as only the range parameter changes while both marginal covariances must be of the same type and have the same sill (if the model is bounded). As this model can be realised as a spatial covariance model with an extra dimension (and anisotropy ratio) to consider the temporal dimension, a number of permissible models already known for spatial covariance models are available for use (Section 2.2.7).

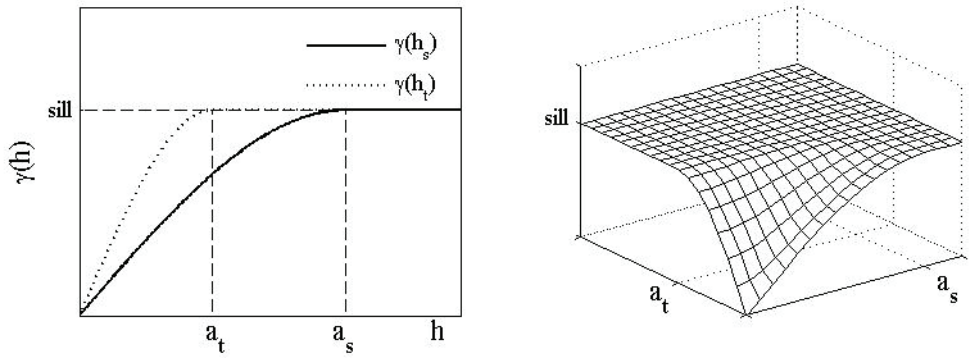


Figure 2.1: Example of metric model, semivariograms in (omnidirectional) spatial and temporal directions (left) and spatiotemporal semivariogram (right).

2.2.8.2 LINEAR MODEL

One type of separability involves adding spatial and temporal covariances (Rouhani & Hall, 1989). This is sometimes referred to as the zonal model due to the analogy with zonal anisotropy and is of the form

$$C_{st}(\mathbf{h}_s, h_t) = C_s(\mathbf{h}_s) + C_t(h_t) \tag{2.43}$$

Using the relationship between $\gamma(\mathbf{h}_s, h_t)$, $C(\mathbf{0},0)$ and $C(\mathbf{h}_s, h_t)$ outlined previously in (2.28), this model can be written in terms of the semivariogram

$$\gamma_{st}(\mathbf{h}_s, h_t) = \gamma_s(\mathbf{h}_s) + \gamma_t(h_t) \tag{2.44}$$

Admissible spatial and temporal covariance models are readily available (outlined later in Section 2.2.7) and are combined to give spatiotemporal models (Figure 2.2). For this model, covariance matrices of certain configurations of spatiotemporal data are singular (Myers & Journel, 1990; Rouhani & Myers, 1990). This means this covariance function is only positive semi-definite and is unsatisfactory for optimal prediction.

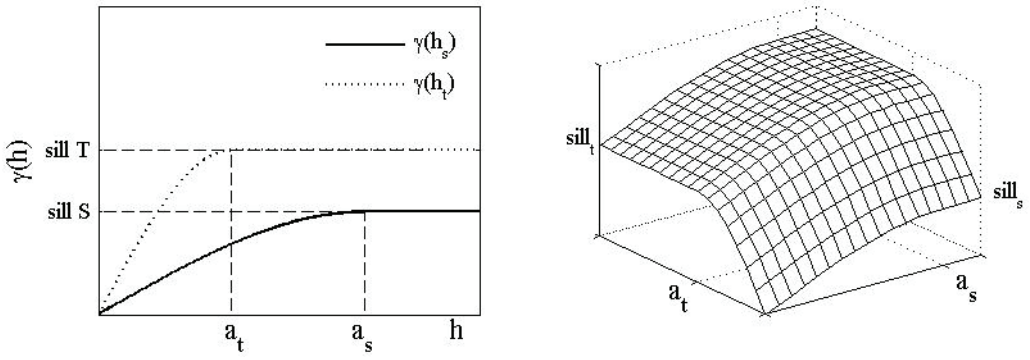


Figure 2.2: Example of linear model, semivariograms in (omnidirectional) spatial and temporal directions (left) and spatiotemporal semivariogram (right).

2.2.8.3 PRODUCT MODEL

The product space-time covariance model (Rodríguez-Iturbe & Mejía, 1974; Posa, 1993; De Cesare et al., 1997) is given by

$$C_{st}(\mathbf{h}_s, h_t) = k \cdot C_s(\mathbf{h}_s) \cdot C_t(h_t) \tag{2.45}$$

where $k \in \mathbb{R}$. This model still separates the spatial dependence from the temporal one. The product covariance model can be rewritten in terms of the spatiotemporal semivariogram:

$$\gamma_{st}(\mathbf{h}_s, h_t) = k\big(C_t(0)\gamma_s(\mathbf{h}_s) + C_s(0)\gamma_t(\mathbf{h}_t) - \gamma_s(\mathbf{h}_s)\gamma_t(h_t)\big) \tag{2.46}$$

The parameter *k* is determined using (2.45) by setting both $\mathbf{h}_s$ and h_t equal to zero.

$$k = \frac{C_{st}(\mathbf{0},0)}{C_s(\mathbf{0}) \cdot C_t(0)} \tag{2.47}$$

It can be shown that this model requires the spatial and temporal sills to be equal to the global sill $C_{st}(\mathbf{0},0)$:

$$\gamma_{st}(\mathbf{h}_s,0) = \frac{C_{st}(\mathbf{0},0)}{C_s(\mathbf{0}) \cdot C_t(0)}\left(C_t(0)C_s(\mathbf{0}) + C_s(\mathbf{0})\gamma_t(0) - C_s(\mathbf{0})\gamma_t(0)\right) = C_{st}(\mathbf{0},0) \tag{2.48}$$

$$\gamma_{st}(\mathbf{0},h_t) = \frac{C_{st}(\mathbf{0},0)}{C_s(\mathbf{0}) \cdot C_t(0)}\left(C_t(0)\gamma_s(\mathbf{0}) + C_s(\mathbf{0})C_t(0) - \gamma_s(\mathbf{0})C_t(0)\right) = C_{st}(\mathbf{0},0) \tag{2.49}$$

Even though this model has the advantage that it is completely determined by the corresponding marginal temporal semivariogram and marginal spatial semivariogram, only one global sill (equal to both the spatial and temporal sill) is possible (Myers, 2004). Whilst this was also the case for the metric model, the product model allows for different types of spatial and covariance models (Figure 2.3). As for the linear and metric models, many spatial and temporal covariance/semivariogram models are readily available. This class of models is severely limited since for any pair of time points, the cross covariance of the two spatial processes always has the same shape (2.50), and similarly for any pair of spatial locations and the cross covariance of the two time series (2.51). (De Cesare et al., 2001a)

$$C_{st}(\mathbf{h}_1,h_t) \propto C_{st}(\mathbf{h}_2,h_t), \text{for fixed spatial lags } \mathbf{h}_1 \text{ and } \mathbf{h}_2 \tag{2.50}$$

$$C_{st}(\mathbf{h}_s,h_1) \propto C_{st}(\mathbf{h}_s,h_2), \text{for fixed temporal lags } h_1 \text{ and } h_2 \tag{2.51}$$

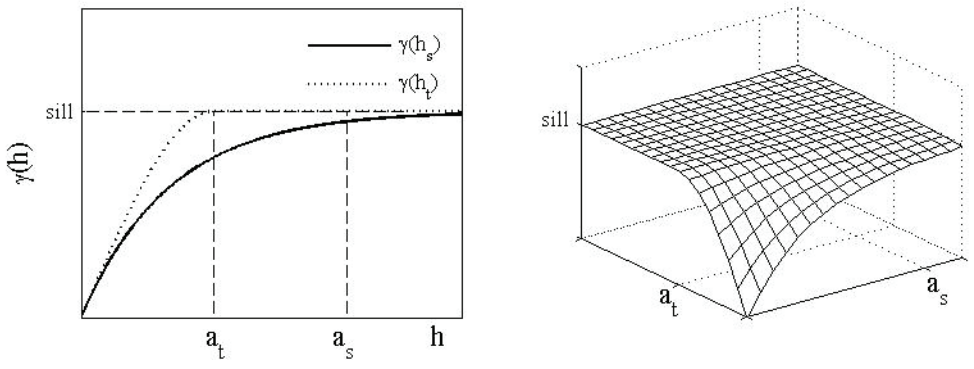


Figure 2.3: Example of product model, marginal (omnidirectional) spatial and temporal semivariograms (left) and spatiotemporal semivariogram (right).

2.2.8.4 PRODUCT-SUM MODEL

The product model can easily be extended to include an additional component involving the product of the spatial and temporal covariance models as follows

$$C_{st}(\mathbf{h}_s,h_t) = k_1 C_s(\mathbf{h}_s)C_t(h_t) + k_2 C_s(\mathbf{h}_s) + k_3 C_t(h_t) \tag{2.52}$$

and can be rewritten in terms of the spatiotemporal semivariogram.

$$\gamma_{st}(\mathbf{h}_s,h_t) = \big(k_1 C_t(0) + k_2\big)\gamma_s(\mathbf{h}_s) + \big(k_1 C_s(\mathbf{0}) + k_3\big)\gamma_t(h_t) \text{-} k_1\gamma_s(\mathbf{h}_s)\gamma_t(h_t) \tag{2.53}$$

The coefficients k_2 and k_3 must be non negative and k_1 must be strictly positive to be an allowable model. From (2.53) we can obtain

$$\gamma_{st}(\mathbf{h}_s,0) = \big(k_1 C_t(0) + k_2\big)\gamma_s(\mathbf{h}_s) = k_s\gamma_s(\mathbf{h}_s) \tag{2.54}$$

$$\gamma_{st}(\mathbf{0},h_t) = \big(k_1 C_s(\mathbf{0}) + k_3\big)\gamma_t(h_t) = k_t\gamma_t(h_t) \tag{2.55}$$

where k_s and k_t can be viewed as coefficients of proportionality between the space-time variograms $\gamma_{st}(\mathbf{h}_s,0)$ and $\gamma_{st}(\mathbf{0},h_t)$ and the spatial and temporal variograms $\gamma_s(\mathbf{h}_s)$ and $\gamma_t(h_t)$ respectively. Setting $\mathbf{h}_s$ and h_t to zero in (2.52) we have

$$C_{st}(\mathbf{0},0) = k_1 C_s(\mathbf{0})C_t(0) + k_2 C_s(\mathbf{0}) + k_3 C_t(0) \tag{2.56}$$

and from (2.54) and (2.55) we obtain

$$k_s = k_1 C_t(0) + k_2, \quad k_t = k_1 C_s(0) + k_3 \tag{2.57}$$

Then we can solve for the coefficients k_1, k_2 and k_3 in terms of the sill values $C_s(\mathbf{0})$, $C_t(0)$, $C_{st}(\mathbf{0},0)$ and the parameters k_s and k_t to obtain

$$k_1 = \frac{k_s C_s(\mathbf{0}) + k_t C_t(0) - C_{st}(\mathbf{0},0)}{C_s(\mathbf{0})C_t(0)},\ k_2 = \frac{C_{st}(\mathbf{0},0) - k_t C_t(0)}{C_s(\mathbf{0})},\ k_3 = \frac{C_{st}(\mathbf{0},0) - k_s C_s(\mathbf{0})}{C_t(0)} \tag{2.58}$$

and in terms of the semivariogram

$$\gamma_{st}(\mathbf{h}_s,h_t) = \gamma_{st}(\mathbf{h}_s,0) + \gamma_{st}(\mathbf{0},h_t) - k\gamma_s(\mathbf{h}_s,0)\gamma_t(\mathbf{0},h_t) \tag{2.59}$$

where

$$k = \frac{k_1}{k_s k_t} = \frac{k_s C_s(\mathbf{0}) + k_t C_t(0) - C_{st}(\mathbf{0},0)}{k_s C_s(\mathbf{0}) k_t C_t(0)} \tag{2.60}$$

Like the product model, the product-sum model has the advantage that it is determined by the corresponding marginal temporal semivariogram and marginal spatial semivariogram, along with the global sill. By estimating and modelling the marginal semivariograms, the spatial and temporal sills $k_s C_s(\mathbf{0})$ and $k_t C_t(0)$ are also obtained. The global sill can be estimated from

the experimental spatiotemporal semivariogram and used, along with the spatial and temporal sills, to calculate *k*. In modelling the separate spatial and temporal semivariograms it is necessary to ensure that the sills are chosen so that the numerators in (2.58) remain positive. A sufficient condition to ensure this is given by (De Iaco et al., 2001)

$$0 < k \leq \frac{1}{\max\left\{sill\left(\gamma_{st}\left(h_s, 0\right)\right); sill\left(\gamma_{st}\left(0, h_t\right)\right)\right\}} \tag{2.61}$$

This requires that the global sill is greater than the individual spatial and temporal sills, but less than their combined value. The product-sum model allows for the specification of different types of covariance/semivariogram models for the spatial and temporal directions (Figure 2.4). It also provides a mechanism for the interaction of space and time, thereby offering more flexibility than the metric, linear and product models.

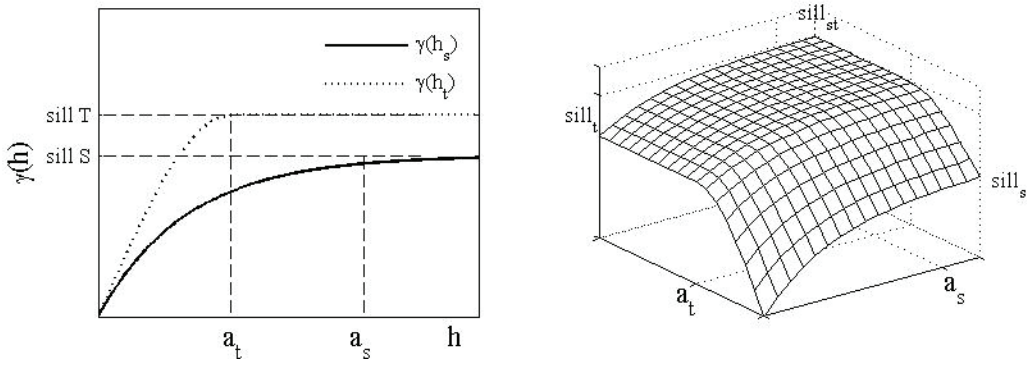


Figure 2.4: Example of product-sum model, marginal (omnidirectional) spatial and temporal semivariograms (left) and spatiotemporal semivariogram (right).

The spatial and temporal marginal structures of the linear, product and product-sum models can be modelled with their own nugget effects. Alternatively, a global nugget effect can be added to all four models presented (De Iaco, 2010).

As for the spatial and temporal semivariogram models, spatiotemporal semivariogram models can be fitted manually or automatically. Statistical criteria, such as weighted least-squares outlined previously (Section 2.2.7.3) can be used to help justify the choice of a particular model and its set of parameters.

There are a number of other spatiotemporal covariance models that will not be considered in the scope of this study. These include, but are not limited to, the Cressie-Huang model (Cressie & Huang, 1999), the Gneiting models (Gneiting, 2002), the integrated product model and mixture models (De Iaco et al., 2002a), the integrated product-sum model (De Iaco et

al., 2002b) and other non-separable spatiotemporal covariance models (Kolovos et al., 2004; Stein, 2005; Porcu et al., 2008; Rodrigues & Diggle, 2010).

2.2.9 SPATIOTEMPORAL ESTIMATION

Once we have a (semivariogram or covariance) model of spatiotemporal dependence we can estimate the attribute value at an unsampled location $(\mathbf{u},t)$. This section presents a family of estimation algorithms called kriging. Kriging is a generic name used for a family of generalised least-squares regression algorithms (Isaaks & Srivastava, 1989; Goovaerts, 1997). All kriging algorithms produce estimates though weighted linear combinations of a subset of the data $\{z(\mathbf{u},t), i=1,...,n\}$, or more specifically, a subset of the residuals $\{z(\mathbf{u},t)-m(\mathbf{u},t)\}$, which are dependent on the specification of the mean function. Data are selected on the basis of spatial and temporal distance from the estimation datum, and weighted by taking into account the proximity of each observation to the prediction location. These spatiotemporal effects are included through reference to the spatiotemporal continuity of the data set modelled by the semivariogram.

Kriging is a best linear unbiased estimator (BLUE) method. It is "linear" since the estimated values are weighted linear combinations of the available data, "unbiased" because the mean of the errors is zero, and "best" since it aims at minimizing the variance of the errors. All kriging estimators are variants of the basic linear regression estimator $Z^*(\mathbf{u},t)$ defined as

$$Z^*(\mathbf{u},t)-m(\mathbf{u},t)=\sum_{\alpha=1}^{n(\mathbf{u},t)}\lambda_\alpha(\mathbf{u},t)\left[Z(\mathbf{u},t)_\alpha-m(\mathbf{u},t)_\alpha\right] \tag{2.62}$$

where $\lambda_\alpha(\mathbf{u},t)$ is the weight assigned to the datum $z(\mathbf{u},t)_\alpha$ interpreted as a realisation of the RV $Z(\mathbf{u},t)_\alpha$ and $m(\mathbf{u},t)$ and $m(\mathbf{u},t)_\alpha$ are the expected values of the RVs $Z(\mathbf{u},t)$ and $Z(\mathbf{u},t)_\alpha$ respectively. The number of data involved in the estimation, along with the weights assigned to each datum, may change from one location to another. In practice, the number of data, $n(\mathbf{u},t)$, closest to the location $m(\mathbf{u},t)$ being estimated are used in the estimation at that location. We call this neighbourhood or window $W(\mathbf{u},t)$ centred on $(\mathbf{u},t)$. It is up to the practitioner to decide the meaning of closest in terms of spatial and temporal distance.

We can define the estimation error as a RV $Z^*(\mathbf{u},t)-Z(\mathbf{u},t)$. The objective of all kriging estimators is to minimise the estimation or error variance $\sigma_E^2(\mathbf{u},t)$ of the estimation error RV under the constraint of the unbiasedness of the estimator. The variance is given by

$$\sigma_E^2(\mathbf{u},t)=Var[Z^*(\mathbf{u},t)-Z(\mathbf{u},t)] \tag{2.63}$$

which is minimised under the constraint that

$$E[Z^*(\mathbf{u},t) - Z(\mathbf{u},t)] = 0 \tag{2.64}$$

The kriging estimator varies depending on the model adopted for the RF $\mathbf{Z}(\mathbf{u},t)$ itself. The RF $\mathbf{Z}(\mathbf{u},t)$ is typically decomposed into a residual component $R(\mathbf{u},t)$ and a trend component $\mathbf{m}(\mathbf{u},t)$

$$\mathbf{Z}(\mathbf{u},t) = \mathbf{m}(\mathbf{u},t) + \mathbf{R}(\mathbf{u},t) \tag{2.65}$$

The residual component is modelled as a stationary RF with zero mean and covariance $C_R(\mathbf{h}_s, h_t)$. The expected value of the RV $\mathbf{Z}(\mathbf{u},t)$ at location $(\mathbf{u},t)$ is the value of the trend component at that location

$$E\big[\mathbf{Z}(\mathbf{u},t)\big] = \mathbf{m}(\mathbf{u},t) \tag{2.66}$$

In simple kriging (SK) we consider the mean $\mathbf{m}(\mathbf{u},t)$ to be known and constant throughout the study area $\mathcal{A}$, whilst in ordinary kriging (OK) we account for local fluctuations of the mean by considering the mean to be constant but unknown in the local neighbourhood.

2.2.9.1 SIMPLE KRIGING

The modelling of the trend component $\mathbf{m}(\mathbf{u},t)$ as a known stationary mean *m* allows us to write the linear estimator (2.62) as a linear combination of the $n(\mathbf{u},t)$ RVs $Z(\mathbf{u},t)_\alpha$ and the mean value *m*

$$\begin{aligned} Z^*_{SK}(\mathbf{u},t) &= \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda^{SK}_\alpha(\mathbf{u},t)[Z(\mathbf{u},t)_\alpha - m] + m \\ &= \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda^{SK}_\alpha(\mathbf{u},t) Z(\mathbf{u},t)_\alpha + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda^{SK}_\alpha(\mathbf{u},t) Z(\mathbf{u},t)_\alpha\right] m \end{aligned} \tag{2.67}$$

SK does not adapt to local trends; rather it relies on a constant, global mean. The mean is subtracted from the observations prior to estimation, and the residual values are used to correct the estimation surface. The $n(\mathbf{u},t)$ weights $\lambda^{SK}_\alpha(\mathbf{u},t)$ are determined such as to minimise the error variance (2.63) under the unbiasedness constraint (2.64) using a system of equations called the SK system given by

$$\sum_{\beta=1}^{n(\mathbf{u},t)} \lambda^{SK}_\beta(\mathbf{u},t) C\big(\mathbf{u}_\alpha - \mathbf{u}_\beta, t_\alpha - t_\beta\big) = C\big(\mathbf{u}_\alpha - \mathbf{u}, t_\alpha - t\big), \quad \alpha = 1,\ldots, n(\mathbf{u},t) \tag{2.68}$$

The weights are determined from the covariance between observations $C(\mathbf{u}_\alpha - \mathbf{u}_\beta, t_\alpha - t_\beta)$ and the covariance between observations and the prediction location $C(\mathbf{u}_\alpha - \mathbf{u}, t_\alpha - t)$. The minimum error variance is called the SK estimation variance and is given by

$$\sigma^2_{SK} = C(0) - \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda_\alpha^{SK}(\mathbf{u},t) C(\mathbf{u}_\alpha - \mathbf{u}, t_\alpha - t) \tag{2.69}$$

2.2.9.2 ORDINARY KRIGING

Ordinary kriging takes account of local variation of the mean by limiting the domain of stationarity of the mean to the local neighbourhood $W(\mathbf{u},t)$. The use of a local mean in OK amounts to re-estimating the mean at each grid node from the data within the search neighborhood. The unknown mean is estimated simultaneously with the residual component. This moving window approach allows some degree of robustness to the assumption of stationarity. The linear estimator (2.62) is then a linear combination of the $n(\mathbf{u},t)$ RVs $Z(\mathbf{u},t)_\alpha$ plus the constant local mean value $m(\mathbf{u},t)$

$$\begin{aligned} Z^*_{OK}(\mathbf{u},t) &= \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda_\alpha(\mathbf{u},t)\left[Z(\mathbf{u},t)_\alpha - m(\mathbf{u},t)\right] + m(\mathbf{u},t) \\ &= \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda_\alpha(\mathbf{u},t) Z(\mathbf{u},t)_\alpha + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda_\alpha(\mathbf{u},t)\right] m(\mathbf{u},t) \end{aligned} \tag{2.70}$$

The unknown local mean $m(\mathbf{u},t)$ is filtered from the estimator by forcing the kriging weights to sum to 1. The OK estimator is then written as a combination of the $n(\mathbf{u},t)$ RVs $Z(\mathbf{u},t)_\alpha$

$$Z^*_{OK}(\mathbf{u},t) = \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda_\alpha^{OK}(\mathbf{u},t) Z(\mathbf{u},t)_\alpha \quad \text{with} \quad \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda_\alpha(\mathbf{u},t) = 1 \tag{2.71}$$

As for SK, the $n(\mathbf{u},t)$ weights $\lambda_\alpha^{OK}(\mathbf{u},t)$ are determined such as to minimise the error variance (2.63) checking for the unbiasedness constraint (2.64) using a system of equations called the OK system given by

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u},t)} \lambda_\beta^{OK}(\mathbf{u},t) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_{OK}(\mathbf{u},t) = C_R(\mathbf{u}_\alpha - \mathbf{u}, t_\alpha - t), \quad \alpha = 1,\ldots,n(\mathbf{u},t) \\ \sum_{\beta=1}^{n(\mathbf{u},t)} \lambda_\beta^{OK}(\mathbf{u},t) = 1 \end{cases} \tag{2.72}$$

where $\mu_{OK}(\mathbf{u},t)$ is a Lagrange parameter that accounts for the constraint on the weights. The minimum error variance is called the OK estimation variance and is given by

$$\sigma_{OK}^2 = C(0) - \sum_{\alpha=1}^{n(\mathbf{u},t)} \lambda_\alpha^{OK}(\mathbf{u},t) C\left(\mathbf{u}_\alpha - \mathbf{u}, t_\alpha - t\right) - \mu_{OK}(\mathbf{u},t) \tag{2.73}$$

Although the mean is assumed stationary only within the local neighbourhood, in practice the residual covariance is assimilated to the global z-covariance inferred from all data available $C_R(\mathbf{u},t) = C(\mathbf{u},t)$. As for the SK system, weights are calculated using the covariance between observations $C\left(\mathbf{u}_\alpha - \mathbf{u}_\beta, t_\alpha - t_\beta\right)$ and the covariance between observations and the prediction location $C\left(\mathbf{u}_\alpha - \mathbf{u}, t_\alpha - t\right)$. The main difference between SK and OK is in the system of equations governing the calculation of the weights. The SK system is modified to obtain the OK system by the addition of a Lagrange multiplier to satisfy the unbiasedness constraint that the weights sum to one.

2.2.10 CROSS VALIDATION

The ultimate objective of the kriging (estimation) methods outlined thus far is typically to estimate the variable at unsampled spatiotemporal locations. It is good practice to validate the entire geostatistical model and kriging plan prior to its implementation to unsampled locations. Cross validation allows us to assess the impact of a semivariogram model and kriging process on interpolation results. It is a measure of the performance of the kriging procedure within the estimation region of the data set. Whilst cross-validation can detect what might go wrong, it does not necessarily ensure that the procedure will be successful. The idea is to remove one datum at a time from the data set and re-estimate this value from the remaining data using the semivariogram model and the kriging procedure outlined in Section 2.2.9.

Interpolated and actual values can be compared to assess the performance of the model by computing the associated errors of the predicted values $e_i = Z_i - Z_i^*$. The distribution of the errors should be symmetric, centred on a zero mean, with a minimum spread. A plot of the errors versus the estimated values should be centred around the zero error line and should have an equal spread. The errors should not show any trend in space or time.

Averaging the quality of the predictions across the validation sets yields an overall measure of prediction accuracy. Typical measures of comparison include:

1. Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Z_i - Z_i^*\right)^2} \tag{2.74}$$

2. Mean error (ME)

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i - Z_i^*\right) \tag{2.75}$$

3. Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Z_i - Z_i^*\right| \tag{2.76}$$

4. Mean percentage error (MPE)

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\frac{Z_i - Z_i^*}{Z_i} \tag{2.77}$$

5. Mean absolute percentage error (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Z_i - Z_i^*}{Z_i}\right| \tag{2.78}$$

6. Correlation coefficient (*r*)

$$r_{ZZ^*} = \frac{1}{(n-1)s_Z s_{Z^*}}\sum_{i=1}^{n}\left(Z_i - \overline{Z}_i\right)\left(Z_i^* - \overline{Z}_i^*\right) \tag{2.79}$$

where *n* is the number of data in the cross-validation set, $\overline{Z}$ is the sample mean of the data, $\overline{Z}^*$ is the sample mean of the estimates, s_Z is the sample standard deviation of the data and s_{Z^*} is the sample standard deviation of the estimates.

RMSE, MAE and MAPE measure the accuracy of the models so the model that yields smaller values for these statistics is assumed to perform better. The ME and MPE measure the bias of the model and should be as close as possible to zero. The correlation coefficient *r* should be as close as possible to 1.

Cross validation has a number of restrictions. Total sills cannot be cross validated from re-estimation scores as rescaling of the semivariogram model does not influence kriging weights, although it does change the kriging variance. The relative nugget effect and semivariogram behaviour at the origin cannot be cross validated as actual values of the semivariogram for lags smaller than the shortest sampling interval are not possible. If the sample data are not representative of the study area, then the model that produces the best cross-validated results may not yield the best predictions at unsampled locations. As many

implementation parameters involving the search strategy and interpolation algorithm can affect re-estimation values, it may be unclear which parameter/s should be changed if the model is inadequate.

2.2.11 JACKNIFE ESTIMATION

In practice, it is possible to reserve a portion of the data set that will not be used in the (semivariogram) modelling process but instead can be used for comparison with estimated values. This is referred to as the validation, or jacknife, data set and can be used to assess the predictive ability of the model in the validation region, which is typically outside the spatial or temporal bounds of the estimation region of the data set. Estimates are made for each element of the validation set and their errors are analysed by techniques outlined previously for the analysis of the cross-validation estimates.

2.2.12 SPATIOTEMPORAL SEQUENTIAL SIMULATION

Whilst geostatistical estimation (kriging) methods focus on deriving an optimal estimate and the associated error variance, geostatistical simulation attempts to model the local uncertainty. From this model, we can derive estimates optimal for different criteria. Different simulation algorithms impart different global statistics and spatiotemporal features on each realisation. No single simulation algorithm is flexible enough to allow the reproduction of the wide variety of features and statistics encountered in practice (Deutsch & Journel, 1998). The sequential simulation algorithm presented here makes use of the kriging estimate and estimation variance obtained at location $(\mathbf{u},t)$ as the parameters of the local distribution.

The objective of simulation is to randomly select a set of *L* equally probable realisations $\{z^{(l)}(\mathbf{u},t)_j;\ j=1,...N, l=1,\ldots,L\}$ from the infinite set of all possible realisations of the RF $\mathbf{Z}(\mathbf{u},t)$. A simulation is said to be conditional if the realisations are generated such that the value of the realisation is equal to the sample value for every sample location. This sequential simulation algorithm is conditional by construction.

The conditional cumulative distribution function (ccdf) of the joint distribution of the set of RVs, from which an *N*-variate sample can be drawn, may be expressed as

$$F\big((\mathbf{u},t)_1,(\mathbf{u},t)_2,...,(\mathbf{u},t)_N;z_1,z_2,...,z_N\mid(n)\big)=\Pr\{Z(\mathbf{u},t)_1<z_1,Z(\mathbf{u},t)_2<z_2,...,Z(\mathbf{u},t)_k<z_k\mid(n)\} \tag{2.80}$$

Sequential simulation algorithms trade the sampling of the multipoint ccdf for the sequential sampling of multiple one-point ccdfs which are easier to infer. Therefore the random drawing

from the *N*-variate ccdf in (2.80) can be carried out by sequential drawing from *N* univariate ccdfs with increasing conditioning (Goovaerts, 1997)

$$F\left((\mathbf{u},t)_1,(\mathbf{u},t)_2,...,(\mathbf{u},t)_N;z_1,z_2,...,z_N\mid(n)\right)=F\left((\mathbf{u},t)_N;z_N\mid(n+N-1)\right)\cdot F\left((\mathbf{u},t)_{N-1};z_{N-1}\mid(n+N-2)\right)...\cdot F\left((\mathbf{u},t)_1;z_1\mid(n)\right) \quad (2.81)$$

N successive steps are taken to generate a realisation, with each step requiring the derivation of the univariate ccdf at each location with an increasing level of conditioning to both sample data and previously simulated values.

All variants of sequential simulation follow the same general procedure with the difference in the algorithms occurring in the step that determines the parameters of the local ccdf. The generalised sequential approach proceeds as follows.

1. Determine a random path that visits each of the *N* nodes only once.
2. Perform the simulation:
    a. At each location estimate (via kriging) the parameters of the local cdf conditional to the sample data and previously simulated values.
    b. Draw the value from the ccdf and add it to the conditioning data.
    c. Proceed to the next node and repeat a) and b). Continue in this manner until all *N* locations have been visited and allocated a simulated value.

Subsequent realisations can be generated by repeating steps 1 and 2. In practice the conditioning data retained in order to determine the ccdf are limited to the sample data and previously simulated values closest to the location being simulated. The statistical properties of the simulation are independent of the visiting sequence of the nodes hence the ordering of the sequence is not important. However in practice it is recommended that each realisation be generated using a random sequence in order to avoid any artefacts which may result from the use of only the closest conditioning data (Chiles & Delfiner, 1999).

2.2.12.1 SEQUENTIAL GAUSSIAN SIMULATION

Realisations of a multivariate Gaussian field can be generated by the sequential algorithm outlined previously. Each variable is simulated sequentially according to its normal ccdf fully characterised through a SK system of type (2.68). If the variable to be simulated is not Gaussian, then the attribute data are transformed using the normal score transform and the

resulting normal score variable is checked for validity of the multiGaussian assumption. In practice, only the biGaussian assumption is checked for normality of the two-point distribution. The sequential Gaussian simulation then proceeds with the generic sequential simulation algorithm steps outlined previously, with the parameters (mean and variance) of the Gaussian ccdf determined at each node using SK with the normal score semivariogram model. The final step is to transform the simulated normal scores into simulated values for the original variable by applying the inverse of the normal score transform. Multiple realisations can be obtained by repeating the entire sequential process with different random paths.

2.2.12.2 POST-PROCESSING OF SIMULATION RESULTS

Simulation output can be interpreted via post-processing of the set of simulated realisations. Kriging gives a smoothed estimate and associated kriging variance at the location being estimated. Using the set of simulated realisations at a particular location (or set of locations), simulation can provide a mean value of the conditional distribution similar to the kriging estimate. In addition, simulation can also provide measures of local variability such as the variance of the conditional distribution, the probability of exceeding a fixed threshold, the value where a particular probability of exceeding that value is reached and measures of the local distribution such as the local histogram.

2.3 GLOBAL TREND MODELLING

Many spatiotemporal data sets have a large-scale (global) trend that cannot be dealt with by the trend component within the geostatistical decomposition into trend and residual component. Large scale trends in the data may cause a problem in estimating the model for the residuals causing statistics such as the semivariogram to become unstable. It is conceptually useful to partition data variation between known measured trends and dependences, and a structured stochastic residual component. This residual component reflects the unknown, unmeasured or simply too complex variables. In such instances, it is customary to remove the global trend before geostatistical analysis, analogous to a data transformation, to reveal the underlying spatiotemporal structure. Large-scale trends can be effectively modelled through process-based knowledge of the variable of interest, or empirically through methods such as trend surface estimation or median polish.

Modelling of the global trend is data-driven. As the data used in this thesis is known historically to display an annual temporal trend, Time Series Modelling of the global trend will

be outlined. Another method incorporating temporal and spatial trend modelling, known as Median Polish Trend Modelling, is also presented.

2.3.1 TIME SERIES MODELLING

Time-series models are based on the analysis of a chronological sequence of observations on a particular variable. The observations may be made yearly, quarterly, monthly, weekly, daily, hourly and so on. Time-series data usually consist of a combination of one or more of the following components

1. Trend is the long-term change in the mean level. It many appear in linear or curvilinear form. If there is no trend, we say that the data are stationary.
2. Seasonal effects are variations whose period is weekly, monthly, annual or related to some other seasonal length.
3. Cyclic changes are variations that occur at a fixed period, typically longer than a year, due to some physical cause other than seasonal effects.
4. Residual or error fluctuations are the erratic movements in a time series that have no set or definable time series pattern.

Various time-series methods are available. In this research we focus on time-series decomposition which assumes that the data can be broken down into the various components and a forecast obtained for each component. This technique is particularly useful to detrend and/or deseasonalise data before they are used in other types of estimation methods. This is the primary focus of time series analysis here.

Time-series decomposition models can be classified into two groups: additive models and multiplicative models. For an additive model, we assume that the data are the sum of their components

$$Y(t) = T(t) + C(t) + S(t) + R(t) \tag{2.82}$$

If the data do not contain one of components, the value for that component is set to zero. In a multiplicative model, the data is the product of the various components

$$Y(t) = T(t) \cdot C(t) \cdot S(t) \cdot R(t) \tag{2.83}$$

and if any of the trend, seasonal variation or cyclic components are missing, then their value is assumed to be one. In this thesis we will be considering the multiplicative time series decomposition model.

Data that are reported daily, weekly, quarterly, monthly, etc and that demonstrate a periodic pattern are said to contain a seasonal component. A seasonal series may or may not be trended and may or may not possess a cyclic component. In most cases seasonality is easier to model than trend or cycle because it has a clearly repeated pattern. Trend may be linear or curvilinear, cycles can be any length and may repeat at irregular intervals, but seasonality is well defined (Gaynor & Kirkpatrick, 1994). Therefore the seasonal component is the first component of the time series that is modelled in the decomposition method. Due to the somewhat irregular nature of a cycle and the amount of data it takes to establish a good pattern, the cyclical component is typically the hardest to model. In many cases the cyclical component is considered as part of the irregular fluctuations in the trend or the time series is assumed to have no cyclical component. As the time series considered in this study are annual, no cyclical component is included in the modelling process, the value of $C(t)$ is assumed to be one, and the model can be written as

$$Y(t) = T(t) \cdot S(t) \cdot R(t) \tag{2.84}$$

2.3.1.1 SPECTRAL ANALYSIS

Spectral analysis is a tool used to detect the presence of a trend or seasonal component in a time series. Any stationary time series can be thought of as the sum of uncorrelated components, each associated with a particular frequency, and the importance of any group of components with frequencies falling into some narrow band is measured by their composite variance. A non-stationary time series $Y(t)$ involving a trend and/or seasonal component is a filtered version of the zero-mean stationary time series $X(t)$. The spectral density function (often shortened to spectrum) of the non-stationary time series $s_Y(\omega)$ is related to the spectrum of the stationary time series $s_X(\omega)$ by

$$s_Y(\omega) = \left|c(e^{i\omega})\right|^2 \cdot s_X(\omega) \tag{2.85}$$

where $\left|c(e^{i\omega})\right|^2$ is called the filter function.

The shape of the spectrum of certain models are known and can help us identify the presence of a seasonal or trend components (Granger & Newbold, 1977):

1. A white noise series has a flat spectrum over the whole range $0 \le \omega \le \pi$.

2. Theoretically a series with a cyclical component of frequency ω_s has a spectrum with an infinitely tall, infinitely narrow, peak having finite area at that frequency. In practice there is a tall, narrow peak at the principle seasonal frequency ω_s and smaller peaks at the secondary seasonal frequencies $j\omega_s$, $j = 2,3,\ldots$.

3. A series with an important trend component will have a strong peak at the very low frequencies as trend has an infinite period.

2.3.1.2 DECOMPOSITION METHOD

With data consisting of a trend component, a single seasonal component of length *L* and an irregular component as shown in (2.84), we can use the following decomposition procedure to model the individual components of a multiplicative time series model.

1. Compute a centred moving average $CMA_L(t)$ of length *L*.

   For *L* odd,

$$CMA_L(t) = \frac{\sum_{j=-(L-1)/2}^{(L-1)/2} Y(t-j)}{L} \tag{2.86}$$

   For *L* even,

$$CMA_L(t) = \frac{\sum_{j=-(L/2-1)}^{L/2} Y(t-j) + \sum_{j=-L/2}^{L/2-1} Y(t-j)}{2L} \tag{2.87}$$

2. Using the $CMA_L(t)$ as an initial estimate of the trend component, divide it into the data to obtain the seasonal component.

$$S(t) \cdot R(t) = \frac{T(t) \cdot S(t) \cdot R(t)}{T(t)} = \frac{Y(t)}{CMA_L(t)} \tag{2.88}$$

3. Assign an index number $\{i(t); i = 1,\ldots,L\}$ to each time instant to indicate the associated season. We choose to arbitrarily assign the index *1* to the season associated with the first time instant.

$$i(t) = t - L\left\lfloor \frac{t-1}{L} \right\rfloor \tag{2.89}$$

4. Remove the irregular component from $S(t) \cdot R(t)$ by computing the average $s(i)$ for each of the L seasons, $i = 1,...,L$. Normalise these to add up to L (number of seasons in a year).

$$s(i) = \overline{S(t) \cdot R(t)}, \quad t = i + kL, \quad k = 0,1,2,..., \quad \hat{s}(i) = \frac{s(i)}{\sum_{i=1}^{L} s(i)} \tag{2.90}$$

5. Deseasonalise the data by dividing by the associated seasonal estimates.

$$d(t) = \frac{Y(t)}{S(t)} = \frac{T(t) \cdot S(t) \cdot R(t)}{S(t)} = T(t) \cdot R(t) \tag{2.91}$$

6. Perform regression analysis on the deseasonalised data to obtain the appropriate trend model (linear, quadratic, exponential, etc).

$$\hat{T}(t) = f[d(t)] \tag{2.92}$$

An estimate for the classical decomposition model at any time instant is then given by

$$\hat{Y}(t) = \hat{T}(t) \cdot \hat{S}(t) \tag{2.93}$$

using the seasonal and trend estimates from (2.90) and (2.92) respectively.

2.3.2 MEDIAN POLISH TREND MODELLING

Many data sets can be interpreted as a response variable measured at varying levels of a number of factors. Median polish is a robust data analysis technique (Tukey, 1977; Hoaglin et al., 1983, 1985, 1986) that uncovers structure in data by examining the impact of its factors to the general statistical model

$$Y(f_1, f_2,..., f_n) = g(f_1, f_2,..., f_n) + R(f_1, f_2,..., f_n) \tag{2.94}$$

where $g(f_1, f_2,..., f_n)$ is some function of the levels of the n factors. The median polish method uses an iterative procedure to obtain a function of the levels of the various factors associated at which the data were recorded and the corresponding residuals

$$\hat{Y}(i, j,..., N) = \hat{g}(i, j,..., N) \tag{2.95}$$

where $i, j,..., N$ are the levels of the corresponding n factors $f_1, f_2,..., f_n$. The trend is not explained externally through exploratory variables but rather by variation of the data itself. This is done via a main effect and several row and column effects.

The use of medians results in a resistant technique where isolated large disturbances in a small number of cells will not affect the common value. Each of the *n* factors has a certain number of values it can have. These are referred to as the levels of a factor. The number of levels does not have to be the same for each factor. Each factor and level combination is represented by a cell in an *n*-way table (the number of cells is the product of the number of levels in each factor).

2.3.2.1 ADDITIVE MODEL

The response variable Y of an *n*-factor model can be modelled by a simple additive model as

$$Y(i, j,...,N) = \mu + \alpha_1(i) + \alpha_2(j) + ... + \alpha_n(N) + R(i, j,...,N) \tag{2.96}$$

where μ is a common mean term, the main effects of the factors $f_1, f_2,..., f_n$ are denoted by $\alpha_1(i), \alpha_2(j),..., \alpha_n(N)$ and vary separately from each other, and *R* is a fluctuation, or residual, term. This simple additive model is also known as the main-effects model and it has a simple interpretation because the separate contributions of the factors are added together.

Median polish trend modelling obtains a fit

$$\hat{Y}(i, j,..., N) = m + a_1(i) + a_2(j) + ... + a_n(N) \tag{2.97}$$

for the model of (2.96) and the corresponding residuals can be calculated as

$$r(i, j,..., N) = Y(i, j,..., N) - \hat{Y}(i, j,..., N) \tag{2.98}$$

An *n*-way table is constructed for the residuals by assigning initial values equal to the original data $Y(i, j..., N)$.

$$r^{(0)}(i, j,..., N) = Y(i, j,..., N) \quad \text{for all } i, j,..., N \tag{2.99}$$

The main effects of each factor are assigned initial values of zero.

$$a_1^{(0)}(i) = a_2^{(0)}(j) = ... = a_n^{(0)}(N) = 0 \quad \text{for all } i, j,..., N \tag{2.100}$$

After assigning initial values, the steps for the iterative procedure of median polish are as follows.

1. Calculate medians for each level of each factor.

$$\begin{aligned} A_1^{(p)}(i) &= \underset{l\neq i}{med}\left\{r^{(p-1)}(i,\, j,...,\, N)\right\} \\ A_2^{(p)}(j) &= \underset{l\neq j}{med}\left\{r^{(p-1)}(i,\, j,...,\, N) - A_1^{(p)}(i)\right\} \\ &\vdots \\ A_n^{(p)}(N) &= \underset{l\neq N}{med}\left\{r^{(p-1)}(i,\, j,...,\, N) - ... - A_2^{(p)}(j) - A_1^{(p)}(i)\right\} \end{aligned} \tag{2.101}$$

2. Subtract medians from the residuals.

$$r^{(p)}(i,\, j,...,N) = r^{(p-1)}(i,\, j,...,N) - A_1^{(p)}(i) - A_2^{(p)}(j) - ... - A_n^{(p)}(N) \tag{2.102}$$

3. Calculate adjustments of effects of this iteration.

$$\begin{aligned} B_1^{(p)}(i) &= A_1^{(p)}(i) - \underset{i}{med}\left\{A_1^{(p)}(i)\right\} \\ B_2^{(p)}(i) &= A_1^{(p)}(i) - \underset{l\neq j}{med}\left\{r^{(p-1)}(i,\, j,...,\, N)\right\} - A_1^{(p)}(i) \\ &\vdots \\ B_n^{(p)}(i) &= A_1^{(p)}(i) - \underset{l\neq N}{med}\left\{r^{(p-1)}(i,\, j,...,\, N)\right\} - ... - A_2^{(p)}(j) - A_1^{(p)}(i) \end{aligned} \tag{2.103}$$

4. Calculate main effect estimates for this iteration.

$$\begin{aligned} a_1^{(p)}(i) &= a_1^{(p-1)}(i) + B_1^{(t)}(i) \\ a_2^{(p)}(j) &= a_2^{(p-1)}(j) + B_2^{(t)}(j) \\ &\vdots \\ a_n^{(p)}(N) &= a_n^{(p-1)}(N) + B_n^{(t)}(N) \end{aligned} \tag{2.104}$$

5. Calculate estimate of common term for this iteration.

$$m^{(p)} = B_1^{(t)}(i) + B_2^{(t)}(j) + ... + B_n^{(t)}(N) \tag{2.105}$$

Repeat the iterative procedure until adjustments of each level $B_1^{(p)}(i), B_2^{(p)}(i),..., B_n^{(p)}(i)$ are (approximately) equal to zero. The errors $r(i,\, j,...,\, N)$ are available at each stage of the iterative procedure or can be calculated for each data point by subtracting the main effect and corresponding row and column effects.

The final estimates given by (2.104) and (2.105) depend somewhat on the order in which the factors are considered. However, this is typically not enough to be of practical importance. If possible, it is useful to check whether a different order of factors results in significantly different results. The iterative procedure typically takes 2 or 3 iterations but can take many more. Sometimes the median polish procedure does not converge as the table has reached a position where no further improvement can be made. In cases where the median

polish procedure does give a result, the fit can be close to being optimal in the least-absolute-residuals sense. However, this is not guaranteed.(Hoaglin et al., 1983)

The median polish method works with both balanced and unbalanced designs. Balanced designs are those in which each cell has the same number of observations and unbalanced designs are those in which the number of observations can vary between cells. This is important as many data sets do not have a value for every factor and level combination, especially for models involving many factors.

2.3.2.2 MULTIPLICATIVE MODEL

The median polish procedure can also be used to model a multiplicative model

$$Y(f_1, f_2,..., f_n) = \mu \cdot \alpha_1(f_1) \cdot \alpha_2(f_2) \cdot ... \cdot \alpha_n(f_n) \cdot R(f_1, f_2,..., f_n) \tag{2.106}$$

by the fit

$$\hat{Y}(f_1, f_2,..., f_n) = m \cdot a_1(f_1) \cdot a_2(f_2) \cdot ... \cdot a_n(f_n) \tag{2.107}$$

A simple multiplicative model requires the use of a logarithmic transformation to enable the use of the method described in 2.3.2.1. The procedure involves taking natural logs of the data and fitting an additive mod el (2.97) to the logarithmic data, which can then be transformed back to the original scale of measurement. This simple approach is applicable when the data involve no zero or negative values (Hoaglin et al., 1985). The parameters of the multiplicative model (2.107) are then obtained by taking exponential functions of the parameters of the additive model (2.97) for $\ln(Y(r,c))$:

$$m = \exp(m_{\ln Y}), a = \exp(a_{\ln Y}), b = \exp(b_{\ln Y}) \tag{2.108}$$

Residuals for the multiplicative model are calculated using

$$R(i, j,..., N) = \frac{Y(i, j,..., N)}{\hat{Y}(i, j,..., N)} \tag{2.109}$$

More complex models include interaction terms, combining additive and multiplicative models, can also be implemented but will not be used in this study.

2.3.2.3 CHOICE OF FACTORS FOR SPATIOTEMPORAL TREND MODEL

The median polish method can be used to extract the spatiotemporal trend of a data set by using the temporal and spatial coordinates as factors. Temporal factors are similar to those involved with a time series analysis. Factors involving the time instants at which the data are recorded

may include year, month, week, day factors. Temporal factors may also include any known seasonal periods such as climate season (i.e. summer, winter) or lunar phase (i.e. full moon, new moon). Spatial components will typically involve the Cartesian coordinates of the data. As mentioned previously, the order in which the iterative algorithm proceeds should be investigated to obtain the most consistent results.

2.3.3 MODEL VERIFICATION AND SELECTION

In the construction of a time series or median polish model to determine the trend of the data, the model must be verified to find out whether the model assumptions are satisfied and if the specified model with its estimated parameters is adequate. Diagnostic verification relies on the analysis of residuals resulting from the identified model. The residual e_i for the *i*th observation in a data set is $e_i = Y_i - \hat{Y}_i$ where Y_i is the *i*th data value and $\hat{Y}_i$ is the corresponding model estimate. The fit describes the data, but incompletely.

Interpolated and actual values can be compared to assess the performance of the model in a similar manner as outlined previously for cross-validation and jacknife estimation. The associated errors of the predicted values are calculated as $e_i = Z_i - Z_i^*$ in the case of an additive model or $e_i = Z_i / Z_i^*$ for a multiplicative model. The distribution of the errors should be symmetric, centred on a zero mean for an additive model (or unity mean for a multiplicative model), with a minimum spread. There should be no obvious pattern in the errors or their variance in space or time.

If two or models pass the diagnostic checking steps then model selection criteria must be employed to select the model which will perform the best. The resulting residuals of different models can be compared using the measures described previously for comparing semivariogram models and kriging algorithms (2.74)-(2.79) to determine which model performs better.

CHAPTER 3

STUDY SITE AND DATA DESCRIPTION

3.1 INTRODUCTION

The space-time data to be modelled are commercial king prawn catch and effort data from the Shark Bay managed prawn fishery in Western Australia. The data were provided by the WA Department of Fisheries and come from voluntary daily logbook information provided by the skippers of the fishing vessels. All fishing boat skippers fill out daily logbooks providing catch (kilograms) by species by grade category and effort (minutes trawled) information for each trawl shot. The start latitude and longitude for each shot are also recorded. The purpose of modelling these data was to produce a model that provided a reliable representation of observed data, thereby providing a greater understanding of the underlying trends in catch rate and its distribution within the fishery in individual annual fishing seasons and between successive annual fishing seasons. In addition, modelling would allow the calculation of reliable estimates of unobserved potential catch rates through extrapolation and prediction, with investigation into how far forecasting was possible.

3.2 SHARK BAY PRAWN MANAGED FISHERY

The Shark Bay Prawn Managed Fishery is located within the waters of Shark Bay off the mid west coast of Western Australia (Figure 3.1). The physical area of the fishery is described as “the waters of the Indian Ocean between latitudes 23°34´ S and 26°30´ S and adjacent to Western Australia on the landward side of the 200 m isobath” (Sporer & Kangas, 2005). The fishery targets western king prawns (*Penaeus latisulcatus*) and brown tiger prawns (*Penaeus esculentus*) and also takes a variety of smaller prawn species including coral prawns and endeavour prawns. King prawns are the dominant species, comprising about 70% of the catch with tiger prawns making up most of the remaining 30%. This study focuses on the king prawns.

Management of the fishery is based on limited entry, crew limitations, gear controls, season and area openings and closures, moon phase closures and daily fishing time controls. These input controls are designed to keep effort at levels that will maintain sufficient spawning stocks and aim towards optimal yields for economic return to industry. Twenty seven boats

were licensed for prawn trawling in the fishery from 2001 to 2004, which is the timeframe focused on in this project. Fishing is undertaken using otter trawls. (Sporer & Kangas, 2005)

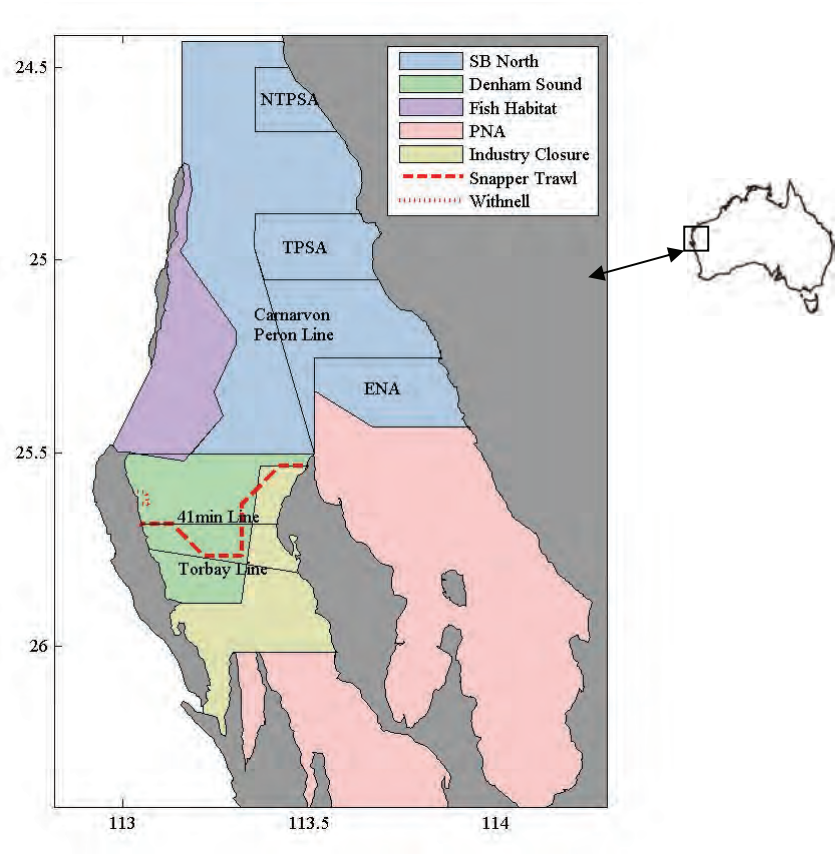


Figure 3.1: Shark Bay Managed Prawn Fishery.

Table 3.1: Prawn season dates, 2001-2004.

	2001	2002	2003	2004
Opening date	14 March	6 March	6 March	16 March
Closing date	28 October	21 October	1 November	25 October

The timing of the opening of the season (Table 3.1) allows the harvest of large residual prawns not caught in the previous year's season. There are a number of permanently closed areas in the fishery including two Permanent Nursery Areas (pink regions, Figure 3.1) to prevent the fishing of small prawns and provide habitat preservation and a Fish Habitat area (purple region) to protect fish species. A Research/Industry Closure Line (yellow region) present in the southern part of Denham Sound in seasons 2001 to 2003 was replaced in the 2004 season by a similar Snapper/Trawl Closure Line (dashed red line). Within the season, there are additional temporary openings and closures including the Carnarvon-Peron Line, Extended Nursery Area (ENA), Tiger Prawn Spawning Area (TPSA), Torbay Line, Denham Line and 41 Minute Line. The opening and closing dates of these lines vary for each season (APPENDIX B).

As the moon phase affects the behaviour of the prawn stocks, there is a three to ten day closure around the full moon phase in each month of the fishing season. Moon phase affects the vulnerability of the prawns at different depths, and through this affects the distribution of fishing (Kangas et al., 2006). These closures aim to increase economic efficiency by shifting fishing effort away from times of reduced catch rate, and reduce the overall level of exploitation on the prawns, thereby assisting to maintain the biomass of females that survive to spawn.

This fishery has two distinct fishing regions, Shark Bay North (blue region, Figure 3.1) and Denham Sound (green region). Shark Bay North involves the fishing region north of the Denham Line whilst Denham Sound includes the Denham Line and the fishing region south of that line. As outlined later, this study focuses on the Shark Bay North fishing region due to the continuity of fishing in that region throughout each season. In contrast, fishing occurs in the Denham Sound region for a period at the beginning and each season, with an extensive period of no fishing in the middle of the season. The periods associated with the two fishing regions are a result of the various closure lines in operation in the fishery throughout the annual season.

3.2.1 FISHING PERIODS

The various closure lines evident in the fishery were implemented at different times in each season creating different fishing periods during each season. The 2001 and 2004 seasons show eight distinct periods whilst those of 2002 and 2003 show six periods. Broadly speaking each of the four fishing seasons (2001 to 2004) can be separated into five similar fishing periods (Figure 3.2) based on the closures line that apply during each period.

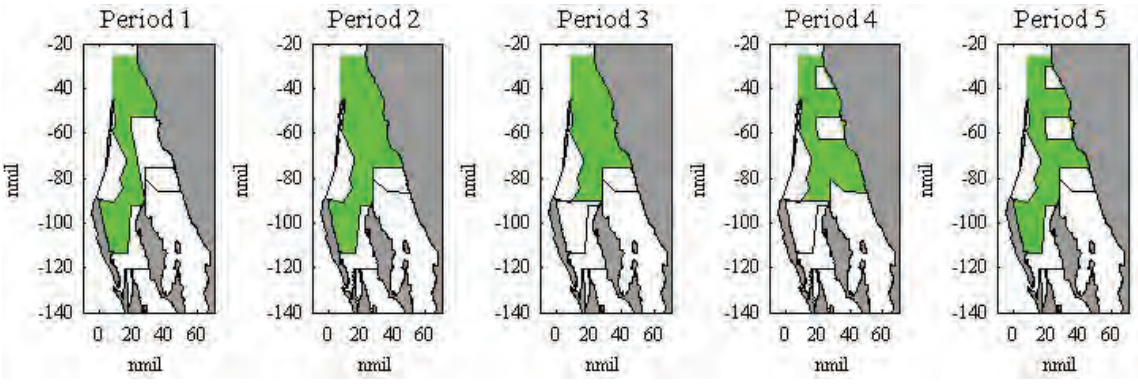


Figure 3.2: Fishing periods evident in the Shark Bay Managed Prawn Fishery, 2001 to 2005. The axes were converted to latitude and longitude in nautical miles (nmil: 1 nmil=1852m) relative to a local coordinate system with origin at latitude -24º and longitude 113º.

The first period of each season beginning in early to mid-March has the Carnarvon-Peron Line and ENA closed and the Denham Line open. The next period begins in mid-April when the Carnarvon-Peron Line opens with the ENA still closed and the Denham Line still

open. The third period marks the closure of the Denham Line in early May with the ENA still closed and the fourth period begins soon after in mid-May when the ENA opens. This period is quite lengthy and includes the closure of the TPSA somewhere from late May to mid-July. The fifth and final period of each season begins in early August when the Denham Line re-opens and the ENA closes. The TPSA is still closed. Each season shows small deviations from these five general fishing periods.

Time series of the weekly average king prawn catch rate in these two areas confirms the discontinuity of fishing in the Denham Sound region during the middle of each season whilst fishing in the Shark Bay North region occurs throughout the season (Figure 3.3). Catch rates are broadly similar in the two fishing regions when they are both open, particularly at the end of the seasons. There are a small number of weeks with no data in either region due to the full moon closures which vary from 3 to 7 days.

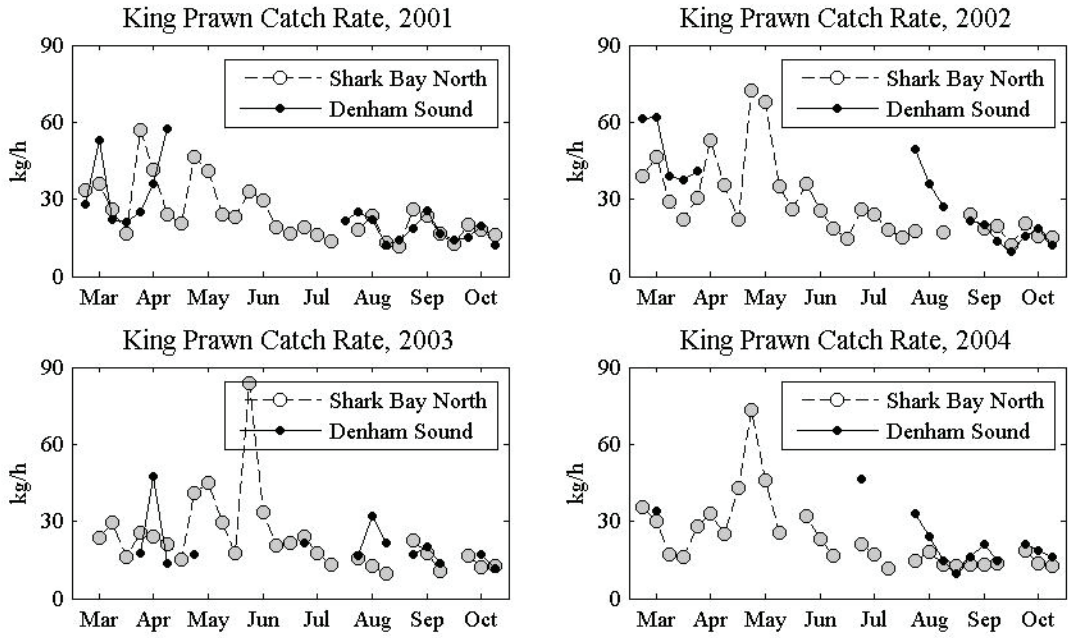


Figure 3.3: Time series of the weekly average king prawn catch rate for the Shark Bay North and Denham Sound fishing regions of the 2001-2004 seasons.

Shark Bay North accounts for approximately 80% of the number of shots and total king prawn catch with Denham Sound making up the remaining 20% (Table 3.2). Due to its continuity within each season and its majority hold of the fishery's catch it was decided to focus this study on the king prawn catch per unit effort (*CPUE*) in the Shark Bay North region of the Shark Bay fishery.

Table 3.2: Number of records and catch within fishery, 2001-2004.

	2001		2002		2003		2004	
Number of shot records	24757		20100		17956		19431	
Shark Bay North	20101	(81.2%)	16145	(80.3%)	15223	(84.7%)	15520	(79.9%)
Denham Sound	4656	(18.8%)	3955	(19.7%)	2745	(15.3%)	3911	(20.1%)
Catch (kg)	1233258		1467446		1109114		1079912	
Shark Bay North	974337	(79.0%)	1173511	(80.0%)	942054	(84.9%)	867094	(80.3%)
Denham Sound	258921	(21.0%)	293936	(20.0%)	167060	(15.1%)	212819	(19.7%)
Effort (hours)	46494		45691		41891		41341	
Shark Bay North	34628	(74.5%)	35430	(77.5%)	33713	(80.5%)	31567	(76.4%)
Denham Sound	11865	(25.5%)	10261	(22.5%)	8178	(19.5%)	9774	(23.6%)
Catch rate (kg/hour)	26.53		32.12		26.48		26.12	
Shark Bay North	28.14		33.12		27.94		27.47	
Denham Sound	21.82		28.65		20.43		21.77	

3.3 SHARK BAY NORTH FISHING REGION

The fishing locations associated with king prawn catch in the Shark Bay North region for seasons 2001 to 2004 are similar between the years (Figure 3.4). The locations do not extend to the boundaries of the fishing region due to fishermen's habits and physical limitations such as seaweed beds. In addition to the logbook catch and effort data for the fishing seasons, recruitment surveys are carried out around the last quarter moon phase in March and April of each season within the Extended Nursery Area and the closed area east of the Carnarvon/Peron line in order to determine the timing and extent of this area to be opened. There are 17 survey sites (Figure 3.5) with catch and effort data for the months of March and April in seasons 2001 to 2004.

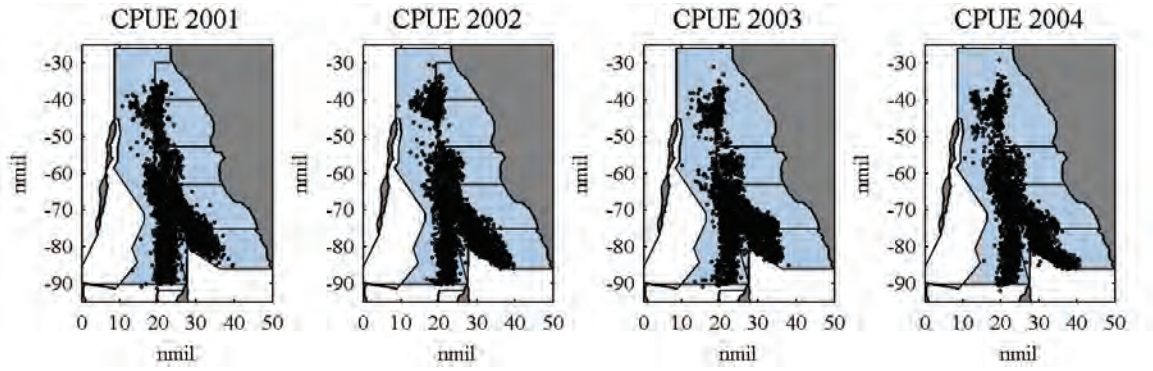


Figure 3.4. Fishing locations in Shark Bay North for seasons 2001 to 2004.

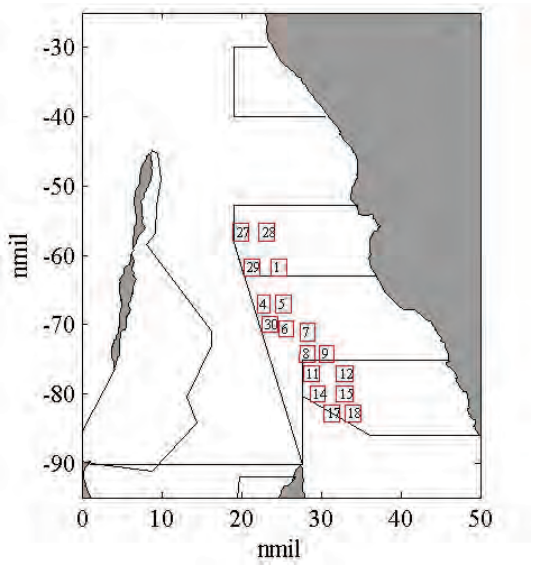


Figure 3.5. Survey site locations in Shark Bay North for seasons 2001 to 2004.

3.4 DATA DESCRIPTION

The data considered for this study are king prawn logbook data recorded by the Shark Bay prawn fishing fleet in the years 2001 to 2004. It is worth noting here that the data preparation was conducted not only for this study but also for a related study concerning the spatial and temporal distribution of king prawns and tiger prawns (Mueller et al., 2008). The data include catch records for the entire prawn fishing seasons for each year (Table 3.1). The original prawn catch data files, obtained from the WA Fisheries and Marine Research Laboratories of the WA Department of Fisheries, included records of the catch, in kilograms, for each species of prawns and for scallops. The file also contained the boat identifier, the data, the shot number of the day, the number of shots made, the latitude and longitude of the start point of each recorded shot and the duration, in minutes, spent fishing.

The data were screened and any records that were missing values for shot location or shot duration were removed from the file. Several records in each year had a missing value in the number of shots column. Rather than removing these records, these values were inferred from the rest of the data, usually as a 1. Records that contained zero tiger or king prawn catch but had a positive catch for either scallops or other prawn species were also removed as it is assumed that neither king nor tiger prawns were being targeted in these shots. This data preparation meant that the data sets for the study contained 24757, 20100, 17956 and 19431 shot records for the years 2001, 2002, 2003 and 2004 respectively (Table 3.3).

Table 3.3: Data preparation, 2001-2004.

	2001	2002	2003	2004
Raw data	26809	20716	19636	21636
Missing Coords/Duration	1529	426	1490	2015
Other species only	523	190	190	190
Shot Data	24757	20101	17956	19431
Shark Bay North	20101	16145	15223	15520
Denham Sound	4656	3955	2745	3911

Shot locations were converted to latitude (*xloc*) and longitude (*yloc*) in nautical miles relative to a local coordinate system with origin at latitude -24º and longitude 113º

$$xloc = 60(longitude - 113)\cos(latitude/180\pi) \tag{3.1}$$

$$yloc = 60\,(latitude + 24) \tag{3.2}$$

3.4.1 WEEK CLASSIFICATION BY LUNAR PHASE

As will be discussed later, the lunar phase is an important indicator of the prawn behaviour and so the data were grouped by the lunar week in which the date of each aggregated record fell. The lunar weeks were calculated by centring each on a phase of the moon. Each observation was designated as occurring during the new moon week (denoted by N), the first quarter week (Q), the full moon week (F) or the last quarter moon week (L) of a particular month of the fishing season. Dates for the moon phases of the 2001-2004 fishing seasons are given in APPENDIX C with the beginning and end dates of each lunar week presented in APPENDIX D. As explained later, data for the full moon weeks was absorbed into the neighbouring first quarter and last quarter weeks.

The order of the weeks within each calendar month with respect to the moon phase is year dependent. For example, the 2001 and 2004 lunar weeks for May ran in the order MayF, MayL, MayN and MayQ. In 2002 the May lunar weeks ran in the order MayL, MayN, MayQ and MayF. In 2003 May ran in the order MayN, MayQ, MayF and MayL.

Since the lunar weeks vary across the seasons it was necessary to choose a method of aligning weeks across successive seasons to enable comparison. It was decided that the optimal alignment would involve the alignment of the 4 moon phases and the approximate alignment of each calendar month across years. Weeks were assigned a number to facilitate modelling. The alignment of weeks for seasons 2001 to 2004, along with their associated week numbers, are shown in APPENDIX E.

As there were three to ten day closures during the full moon weeks, it became apparent that most of the full moon weeks comprised data from the end of the previous week (First Quarter) and data from the beginning of the following week (Last Quarter). Therefore, it was decided to absorb the data for the full moon weeks into the preceding First Quarter week and the following Last Quarter week. This resulted in data sets with data associated with only three of the four moon phases. The week numbers with data and their associated lunar month and phase are shown in Table 3.4. The Last Quarter moon weeks of 2003 are associated with the previous month than those of the other seasons.

Table 3.4: Week numbers and associated lunar month and phase, 2001-2004.

	2001, 2002, 2004		2003	
Week	Month	Phase	Month	Phase
1	Mar	L		
2	Mar	N	Mar	N
3	Mar	Q	Mar	Q
5	Apr	L	Mar	L
6	Apr	N	Apr	N
7	Apr	Q	Apr	Q
9	May	L	Apr	L
10	May	N	May	N
11	May	Q	May	Q
13	Jun	L	May	L
14	Jun	N	Jun	N
15	Jun	Q	Jun	Q
17	Jul	L	Jun	L
18	Jul	N	Jul	N
19	Jul	Q	Jul	Q
21	Aug	L	Jul	L
22	Aug	N	Aug	N
23	Aug	Q	Aug	Q
25	Sep	L	Aug	L
26	Sep	N	Sep	N
27	Sep	Q	Sep	Q
29	Oct	L	Sep	L
30	Oct	N	Oct	N
31	Oct	Q	Oct	Q

Due to the various closure lines operating in the fishery, the data were screened to remove those data at locations that were in closed areas. To do this, each season was split into periods of fishing according to the closure lines and their dates.

3.4.2 DATA AGGREGATION

In many cases catches were recorded in aggregated form over several shots rather than by individual trawl shot. These aggregated records account for a large proportion of the total king prawn catch for each season in Shark Bay North, and the entire fishery, from 41.2% in 2001 to 59.5% in 2002 (Figure 3.6). It was decided to use data aggregated by vessel for analysis by aggregating the individual shots of each vessel on each day for direct use with the (already) aggregated data. A second form of aggregation where data were aggregated by grid location was also considered, resulting in a second data set for analysis. It is noted that this method combines individual shot data and aggregated shot data.

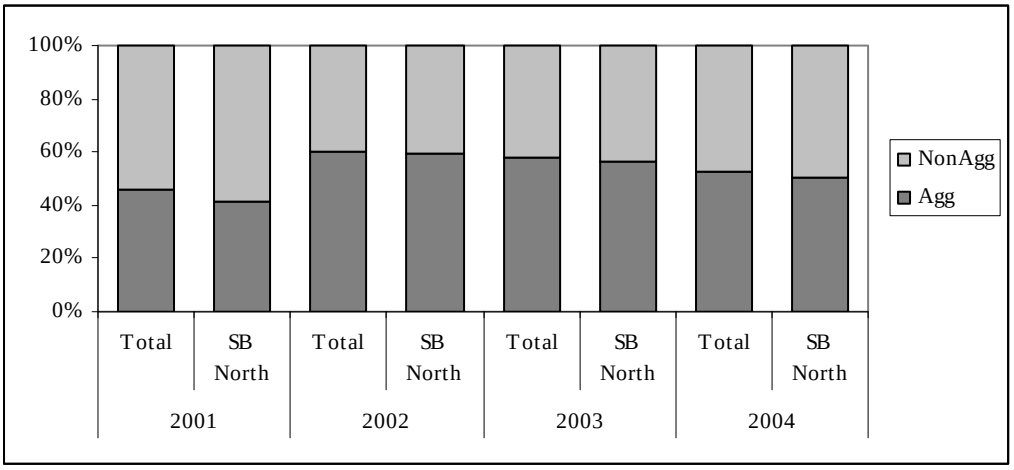


Figure 3.6: Proportion of aggregated king prawn catch by weight for the 2001-2004 seasons.

3.4.2.1 AGGREGATION BY VESSEL

Centroidal locations were calculated for a particular vessel *v* on a particular day *t* by weighting the coordinates of each shot of the day by its duration,

$$x_v(t) = \frac{\sum_{i=1}^{n_v(t)} \left[d_v^i(t)\, xloc_v^i(t) \right]}{\sum_{i=1}^{n(v)} d_v^i(t)}, \quad y_v(t) = \frac{\sum_{i=1}^{n_v(t)} \left[d_v^i(t)\, yloc_v^i(t) \right]}{\sum_{i=1}^{n(v)} d_v^i(t)} \tag{3.3}$$

where $n_v(t)$ denotes the number of shots for that vessel on day *t*, $d_v^i(t)$ the duration of the *i*th shot for that vessel and $xloc_v^i(t)$ and $yloc_v^i(t)$ the corresponding latitude and longitude of each shot in nautical miles. Aggregate catch and effort values were calculated for the new centroidal locations by combining the catch and effort values respectively for all shots of that vessel on the given day *t*. The catch (kg) and effort (mins) for a particular vessel *v* on a particular day *t* are defined as

$$CATCHvd[x_v(t), y_v(t), t] = \sum_{i=1}^{n_v(t)} c_v^i(t) \quad , \quad EFFvd[x_v(t), y_v(t), t] = \sum_{i=1}^{n_v(t)} d_v^i(t) \tag{3.4}$$

where $c_v^i(t)$ denotes the catch associated with the *i*th shot of that vessel on day *t*. The *vd* in the variable names indicates that the data are aggregated by *vessel* on a *daily* basis.

As the kriging and simulation procedures to be used later do not allow more than one observation to be given at a particular location it was necessary to screen the data to ensure that no two or more records had identical spatial locations on the same day. This screening process was performed for locations during the same week for consistency with an associated study (Mueller et al., 2008). Records with identical locations and lunar weeks had their longitudinal location altered by at most ± 0.01 nautical miles to enable their use in the analysis.

3.4.2.2 AGGREGATION BY GRID LOCATION

The chosen grid size for aggregating the shot data was one nautical mile. The origin of the grid was (10,-90) for the local coordinate system, specified previously in Equations (3.1)-(3.2), with 31 nodes in the x-direction and 61 nodes in the y-direction. Aggregate catch and effort values were then calculated at each grid node for each day of the season for the 1 nmil square specified by a set of four grid nodes. The catch (kg) and effort (mins) for a grid node *g* centred on (x_g, y_g) on a particular day are defined as

$$CATCHgd(x_g, y_g, t) = \sum_{i=1}^{n_g(t)} c_g^i(t) \quad , \quad EFFgd(x_g, y_g, t) = \sum_{i=1}^{n_g(t)} d_g^i(t) \tag{3.5}$$

where $n_g(t)$ denotes the number of shots in grid *g* on day *t*, and $c_g^i(t)$ and $d_g^i(t)$ are the catch and duration associated with the *i*th shot in the grid *g* on that day. The *gd* in the variable names indicates that the data are aggregated *daily* by *grid* location.

3.4.2.3 AGGREGATION BY LUNAR WEEK

Weekly catch and effort variables aggregated over the entire study region were also produced for use in time series analysis by aggregating catch and effort values respectively for all shots from all vessels in a lunar week.

$$CATCHw(w) = \sum_{i=1}^{n_t(w)} \sum_{j=1}^{n(t)} c_w^j(i) \quad , \quad EFFw(w_t) = \sum_{i=1}^{n_t(w)} \sum_{j=1}^{n(t)} d_w^j(i) \tag{3.6}$$

where $n_t(w)$ is the number of days in the lunar week *w* and $n(t)$ is the number of individual shots on day *t*. The catch and effort of the *j*th shot of the *i*th day in lunar week *w* are denoted by

$c_w^j(i)$ and $d_w^j(i)$ respectively. Each day was assigned to a lunar week as outlined previously (Section 3.4.1).

3.4.3 CATCH PER UNIT EFFORT

As sampling effort may differ from area to area, month to month, or year to year; the number of fish captured must be analysed in such a way as to standardise the effort that was exerted. Catch per unit effort (CPUE), or catch rate as mentioned previously, standardises catch data based on the amount of the effort (total time or area sampled) exerted. It is often used as a means of tracking the performance of a fishery over successive seasons. In some fisheries CPUE is used as a measure of relative abundance, however there are many problems in using them as such, particularly as a measure of local abundance (Harley et al., 2001; Walters, 2003; Maunder et al., 2006). The relationship between catch rates and stock abundance is given by

$$CPUE = q\beta \tag{3.7}$$

where β is the biomass of the population and q is a fixed constant of proportionality known as the catchability coefficient (Hilborn & Walters, 1992).

Raw CPUE is seldom proportional to abundance over a whole exploitation history and an entire geographic range, because numerous factors affect catch rates. One of the most commonly applied fisheries analyses is standardisation of CPUE data to remove the effect of factors that bias CPUE as an index of abundance (Campbell, 2004). CPUE can also be adjusted for relative efficiency, which changes over time, by calculating and applying relative fishing powers. In this study, we assume that all vessels have the same efficiency and that this has not changed over the four years of study. Even if CPUE is standardised appropriately, the resulting index of relative abundance can be problematic as an indicator of abundance (Bishop et al., 2008). Therefore, CPUE-based analyses are not an alternative to integrated stock assessments, but they provide a component of information for an integrated model. Annual CPUE indices are used as conditioning data in an integrated model, and when used as such, they typically drive the model. In addition, CPUE indices for a fishery, or region within a fishery, provide useful information about fisheries changes, aspects for discussion and indicate areas of future research. In this way, CPUE trends are one of the indicators of the general ‘health’ of the fishery. CPUE is a fishery-dependent indicator, meaning that the data used to calculate the metric are taken directly from the fishing industry.

CPUE is defined as the total catch divided by the total amount of effort used to harvest the catch. As each shot record of the prawn data has an associated catch (kg) and effort (mins) values, the CPUE (kg/h) can be calculated.

$$CPUE = \frac{catch}{effort/60} \tag{3.8}$$

Note the division of effort by 60 as effort is recorded in minutes whilst CPUE is specified per hour. Catch rate variables were calculated for each of the aggregated data sets outlined previously; aggregation daily by vessel, aggregation daily by grid and aggregation by lunar week over entire fishing region.

- ***CPUEvd: Daily aggregation by vessel***

$$CPUEvd[x_v(t), y_v(t), t] = 60 \frac{\sum_{i=1}^{n_v(t)} c_v^i(t)}{\sum_{i=1}^{n_v(t)} d_v^i(t)} \tag{3.9}$$

- ***CPUEgd: Daily aggregation by grid***

$$CPUEgd(x_g, y_g, t) = 60 \frac{\sum_{i=1}^{n_g(t)} c_g^i(t)}{\sum_{i=1}^{n_g(t)} d_g^i(t)} \tag{3.10}$$

- ***CPUEw: Aggregation by lunar week over entire fishing region***

$$CPUEw(w) = 60 \frac{\sum_{i=1}^{n_t(w)} \sum_{j=1}^{n(t)} c_w^j(i)}{\sum_{i=1}^{n_t(w)} \sum_{j=1}^{n(t)} d_w^j(i)} \tag{3.11}$$

3.4.4 STATISTICAL ANALYSIS

The spatiotemporal catch rate variables *CPUEvd* and *CPUEgd* are the main focus of this investigation, with the *CPUEw* variable to be used for temporal trend analysis. There are approximately double the number of *CPUEgd* records as *CPUEvd* records for each season with about 6000-8000 aggregated by 1 nmil grid location (*CPUEgd*) records in the Shark Bay North fishing region for all four seasons but only 3000 aggregated by vessel (*CPUEvd*) records for each of the seasons (Table 3.5). This 2:1 ratio between the *CPUEvd* and *CPUEgd* data is also evident in the individual bins of the histogram (Figure 3.7). All seasons show similar means and

medians with the *CPUEgd* variables showing slightly higher means and medians than the corresponding *CPUEvd* variables. Both methods of aggregation display similar central measures.

Table 3.5. Summary statistics of daily king prawn CPUE (kg/h) aggregated by vessel or grid location.

	2001		2002		2003		2004	
	vd	gd	vd	gd	vd	gd	vd	gd
Mean	27.97	30.61	34.04	36.44	28.00	29.61	27.34	29.94
Median	23.68	25.62	26.65	28.94	21.60	23.03	22.06	24.27
Variance	298.21	411.90	589.72	719.47	533.15	582.21	407.69	509.98
Skewness	1.72	1.77	1.92	2.17	2.80	2.85	1.78	2.32
Min	1.07	1.28	0.92	1.57	1.08	1.08	0.95	0.72
Max	146.07	184.00	210.90	376.70	212.94	330.71	157.21	265.29
Count	3344	7981	3274	7205	3169	6422	2928	6116

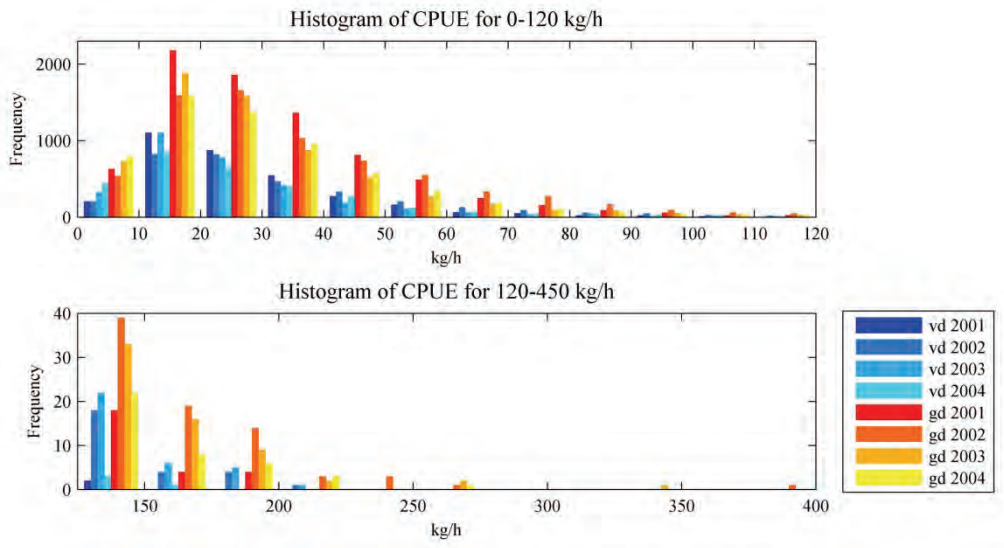


Figure 3.7: Histograms of daily king prawn CPUE (kg/h) aggregated by vessel or grid location for 0-125 kg/h (top) and 125-550 kg/h (bottom).

The minima of all data sets are close to zero with the *CPUEgd* data showing slighter higher minima than the corresponding *CPUEvd* data. The maxima of the *CPUEgd* data (265-526) are considerably higher than the corresponding *CPUEvd* data (146-213). The higher maxima for the *CPUEgd* data are visible in the histograms for the CPUE data. It is noted that there are only two *CPUEvd* values above 200kg/h (210.90 in 2002, 212.94 in 2003) whilst there are 17 *CPUEgd* values above that value ranging from 204.61 to 376.70 (Table 3.5). These outliers have been checked and whilst their values are suspect, they cannot be justified as erroneous. However, it is possible that these extreme values in the *CPUEgd* data may be a result of fishers incorrectly recording catch or effort to a particular shot and therefore possibly erroneous. It was decided to leave these values in the analysis.

All seasons display high variances with the 2002 season showing the highest variances for both types of catch rate data. The variances of the *CPUEgd* variables are higher than the corresponding *CPUEvd* variables. All four seasons of data for both methods of aggregation are positively skewed (Figure 3.7). The *CPUEgd* data sets show slightly higher values of skewness than the corresponding *CPUEvd* data sets due to the higher maxima outlined previously.

Weekly catch rate values (*CPUEw*) will be used to model temporal trends for the individual seasons (Figure 3.8). As outlined previously the data are grouped into lunar weeks based on the moon phase, with data in weeks corresponding to the Full Moon reallocated to the neighbouring Last Quarter and First Quarter weeks.

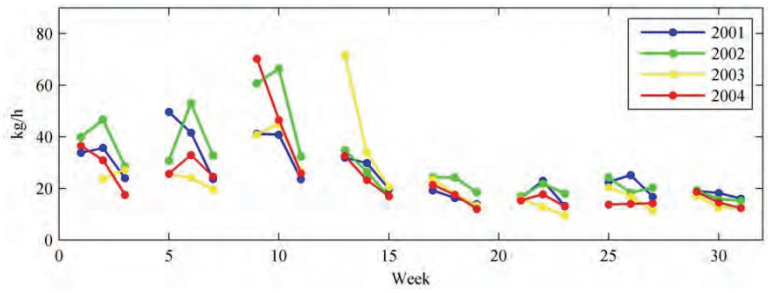


Figure 3.8: Weekly king prawn CPUE (kg/h) aggregated over Shark Bay North region.

The *CPUEw* data show similar means across all seasons with medians slightly lower than the corresponding mean (Table 3.6). The means and medians of the weekly data are slightly lower than the daily catch rate data, aggregated by vessel or grid, of the corresponding season. The *CPUEw* are also positively skewed but show slightly lower values of skewness than the corresponding *CPUEvd* and *CPUEgd* data. The minima/ maxima of the weekly catch rate data sets are much higher/lower than the corresponding daily data sets with minimum values ranging from 9.31 to 15.01, and maximum values between 49.19 and 71.32.

Table 3.6. Summary statistics of weekly king prawn catch rate (kg/h).

	2001	2002	2003	2004
Mean	25.30	28.99	23.34	23.23
Median	22.86	23.97	20.19	18.02
Variance	103.38	206.60	192.64	177.76
Skewness	0.86	1.34	2.15	2.14
Min	12.85	15.01	9.31	11.64
Max	49.19	65.78	71.32	69.67
Count	24	24	23	24

3.4.5 OVERVIEW OF SPATIAL DISTRIBUTION

The daily king prawn CPUE aggregated by vessel and grid demonstrate similar spatial behaviour across the season, which is driven mainly by the closure lines and by the main migration pattern of prawns and the way they move from the south eastern parts of the bay in a north, north westerly direction. The lunar month of May (Figure 3.9, aggregated by vessel, and Figure 3.10, aggregated by grid) is chosen for illustrative purposes throughout this thesis as it highlights the opening of a fishing region (ENA) resulting in high catch rates concentrated just south of the closure boundary. Similarly there are high catch rates to the east of the Carnarvon-Peron line as it opens to fishing in April. The catch rates remain at a reasonable level up until June in Shark Bay North, but are at relatively low levels across the fishery by the end of each season. The lunar month of July (Figure 3.9, aggregated by vessel, and Figure 3.10, aggregated by grid) is also chosen for illustrative purposes throughout this thesis as it is indicative of catch rates in the second half of the season. Regional trends of all four seasons show similarities of high and low catch rate areas throughout the year (APPENDIX F). There is an obvious decline in catch rates seen during each season.

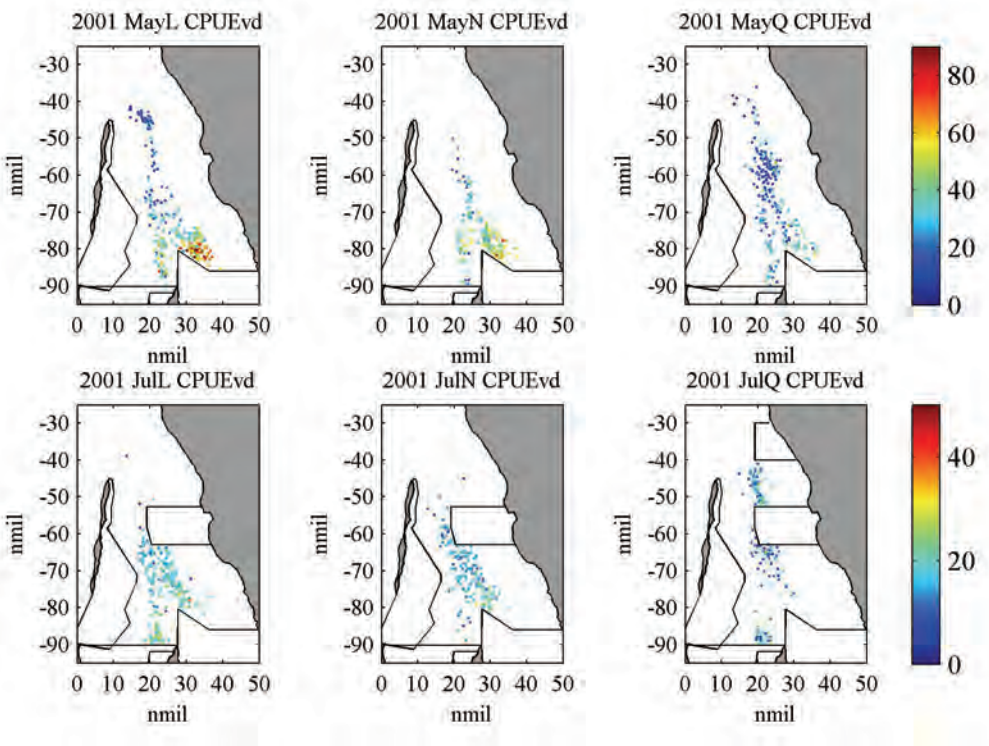


Figure 3.9: Weekly king prawn CPUE (kg/h), May (top) and July (bottom) 2001, aggregated by vessel.

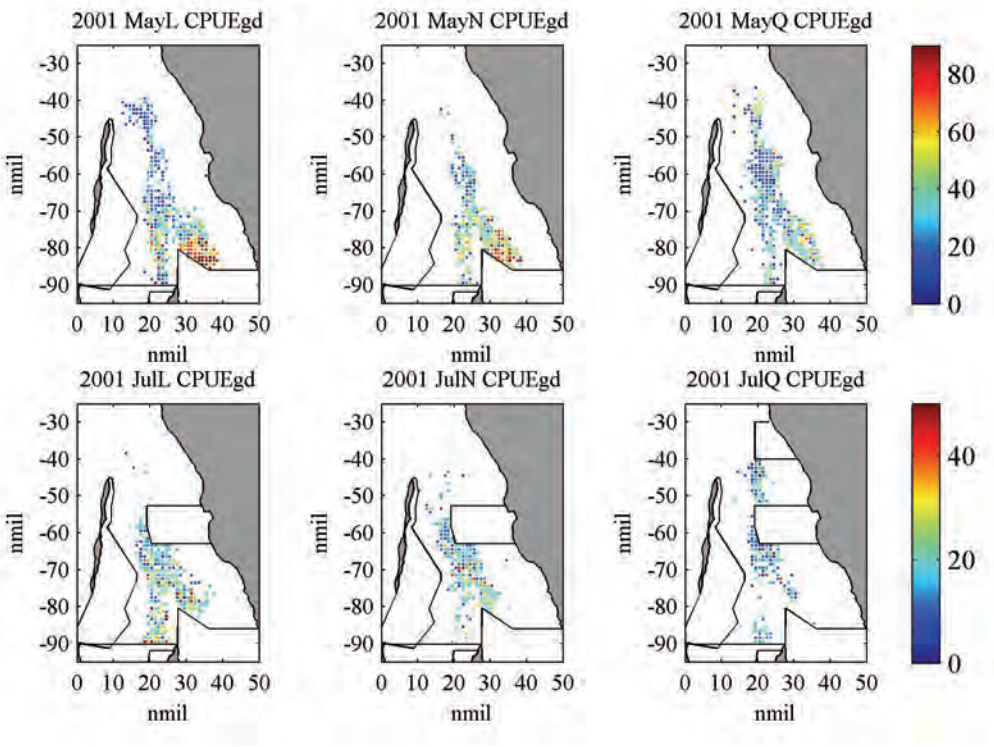


Figure 3.10: Weekly king prawn CPUE (kg/h), May (top) and July (bottom) 2001, aggregated by grid.

3.4.6 SURVEY DATA

As outlined in Section 3.3, there were seventeen fishery-independent recruitment survey locations across the study region sampled around the last quarter moon phase in the months of March and April of seasons 2001 to 2004. These seventeen locations are situated within the Extended Nursery Area and the closed area east of the Carnarvon/Peron line (Figure 3.11, right). The information from the recruitment surveys was used to determine the timing and extent of areas to be opened. The catch rates at a given survey location are quite similar for the March and April surveys (Figure 3.11, left). The northern survey locations (Sites 27-29,1) have the lowest catch rates with higher catch rates observed at the central sites in the survey region (Sites 4-9, 30). Survey sites in the Extended Nursery Area (Sites 11, 12, 14, 15, 17 and 18) show the greatest variability.

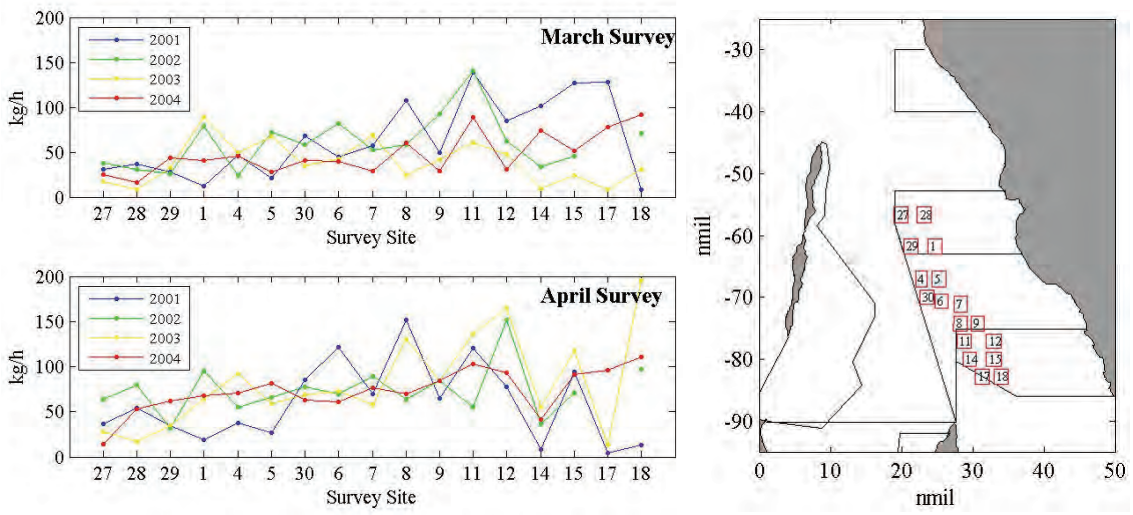


Figure 3.11. Recruitment survey catch rates in March (top left) and April (bottom left) and survey locations (right) in seasons 2001-2004.

3.4.7 PRELIMINARY INVESTIGATION OF SUITABLE SPATIOTEMPORAL GEOSTATISTICAL MODEL

Preliminary investigation of a suitable geostatistical spatiotemporal model for the king prawn catch rates considered a spatially correlated time series framework. Annual time series of daily catch rates for grids of varying mesh size defined over the study region were constructed. However, there were a large number of blocks with missing values for many days over each season. Subsequently, daily data were aggregated to weekly data in an attempt to minimise the number of blocks with missing values but this was unsuccessful.

An alternative model of temporally correlated random functions was also investigated to build a linear model of coregionalisation using weekly catch rate data to limit the number of time instants. The catch rate data were aggregated by grid to ensure data locations were collocated at various weeks. Cross-variography of subsequent weeks did not reveal sufficient evidence of structure for semivariogram modelling.

Spatial variography of (lunar) weekly sets of the daily king prawn catch rates had previously shown that the catch rates displayed spatial continuity (Mueller et al., 2008). Semivariograms of the king prawn *CPUE* were modelled with spatial variograms involving a nugget effect and one or two spherical structures. The behaviour of the range and nugget parameters over a season and between seasons were considered in an effort to expose a pattern of semivariogram model parameters that could be used to infer parameters for a future season. The spatial variogram parameters did show some similarities across weeks of a season but no evidence of temporal pattern was observed.

After consideration of the geostatistical spatiotemporal models outlined above, it was decided to proceed with the single spatiotemporal random function model with the prawn catch and effort data. This model allowed the spatial continuity of the catch rates modelled previously to be extended to include a temporal dimension. By using input data within the temporal range of the model, estimates of catch rates in future weeks was possible.

CHAPTER 4

METHODOLOGY OF THIS STUDY

4.1 INTRODUCTION

Spatiotemporal models are usually data driven and typically constructed by several components. Data preparation has provided two alternative catch rate variables for use: *CPUEvd* is the catch rate per day for a particular vessel (27 vessels) and *CPUEgd* is the catch rate per day on a grid over the study region. In this chapter an outline of the modelling framework for this study is provided including the modelling of the global trend, the modelling of spatiotemporal semivariograms, the framework for forward estimation and simulation of catch rates, and the software used to conduct these analyses.

4.2 GLOBAL TREND MODELLING

As outlined previously, the use of the random function model requires spatiotemporal (intrinsic) stationarity which is not a property of the variable itself, and cannot be proven or refuted by data. It is the modeller's decision of the random function model and cannot be statistically tested or objectively rejected or accepted. Analysis of the catch rate data showed the presence of a large scale temporal trend (Figure 3.3). Geostatistical models aim to include the trend as part of the modelling process. However, large-scale trends are best often removed before modelling of the spatiotemporal continuity by geostatistical methods. The trend is not a smoothly varying function of the spatiotemporal coordinates (x, y, t) for use in a Universal Kriging model (Goovaerts, 1997) nor is it a linear function of an external variable for use in a Kriging with an External Drift model (Goovaerts, 1997). Furthermore, the trend model of the catch rates was identified as multiplicative whilst the random function model assumes an additive trend.

Therefore, it was decided to utilise a global trend model to detrend and deseasonalise the catch rate data in order to obtain the *CPUEs* (standardised catch rate) variable. By identifying and removing the large-scale multiplicative temporal trend, the resulting *CPUEs* variable would be seen as small-scale variation involving autocorrelated departures from the local mean. Decomposition of the data into a large-scale deterministic trend component and residual small-scale component of variation allows the assumption of stationarity to be satisfied. This supports the use of the random function model and modelling by geostatistical methods.

4.2.1 TEMPORAL TREND MODELLING

Temporal trends of prawn catch rates in the Shark Bay managed prawn fishery had been previously analysed by time series analysis using classical decomposition (Harman, 2001). Harman modelled daily king prawn and tiger prawn catch rate values with a multiplicative classical decomposition model with lunar phase identified as the seasonal factor. The centred moving average was used as an estimation of the trend component in conjunction with seasonal indices calculated on a 29 day cycle.

As individual fishing days cannot be compared across seasons, this study considered (lunar) weekly data to enable comparison. Annual king prawn catch rate time series of weekly data (*CPUEw*) were modelled via a multiplicative classical decomposition model using the lunar phases as seasons. Temporal trends of smaller blocks within the Shark Bay North region were also investigated but revealed that modelling the temporal trend on a smaller spatial scale was not feasible. As temporal trend models across the individual seasons showed great similarities, it was also possible to create a model involved combined consecutive seasons to use in estimating the trend model for a future season. Consecutive seasons of data could be combined for analysis by aligning the lunar weeks of the seasons. Temporal trend models for the six combined seasons (involving consecutive seasons) 2001-2002, 2001-2003, 2001-2004, 2002-2003, 2002-2004 and 2003-2004 were considered.

4.2.2 MEDIAN POLISH TREND MODELLING

Median polish trend modelling has been used for estimation of a space-time trend in a geostatistical modelling framework (Bruno, 2004). A multiplicative median polish spatiotemporal model was considered for the king prawn catch rate data with two temporal effects (lunar month and phase) and two spatial effects (longitude and latitude converted to x and y coordinates in nautical miles). The median polish model confirmed the temporal trend seen in the classical decomposition model but showed no significant difference in the resulting standardised catch rate variable over the temporal trend model.

4.2.3 STANDARDISED CATCH RATE

The objective of the methods of global trend modelling (time series and median polish) is to obtain a good fit for the trend of the catch rate data which can be assessed by analysing the residuals from the fitted model. These methods are used to identify the global trend of the spatiotemporal data, which can then be used to transform (detrend and deseasonalise) the data to obtain the standardised catch per unit effort (*CPUEs*) for use in the subsequent geostatistical

model. Both the *CPUEvd* and *CPUEgd* variables were transformed to the corresponding *CPUEsvd* and *CPUEsgd* variables to examine their use in a spatiotemporal geostatistical model.

4.3 GEOSTATISTICAL MODELLING

The standardised catch rate variables (*CPUEsvd* and *CPUEsgd*) obtained from the original catch and effort data by aggregation and global detrending methods (Figure 4.1) were used to develop a modelling framework to implement geostatistical prediction of the king prawn catch rate for a future season.

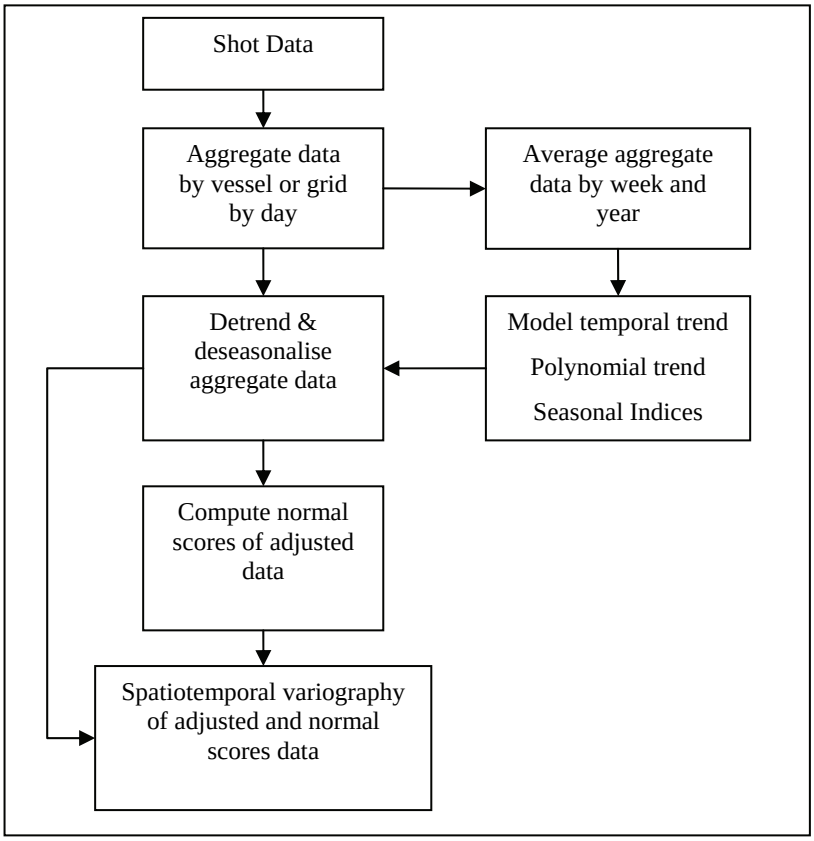


Figure 4.1. Flow chart for temporal trend modelling and spatiotemporal variography.

4.3.1 SPATIAL VARIOGRAPHY

Spatial variography of weekly data sets of king prawn catch rate values was undertaken in a previous study (Mueller et al., 2008). This was repeated for the *CPUEsvd* and *CPUEsgd* and their normal scores (*CPUEsNvd* and *CPUEsNgd*) in this study and confirms the use of geostatistical methods with the standardised prawn catch rate values.

Preliminary analysis considered the approach of temporally correlated (spatial) random functions. Whilst the number of days in each season was too large to implement this method, it was hoped that it was possible by considering weekly aggregated data. However, spatial random functions of weekly data revealed cross variography that proved problematic due to poor correlation.

4.3.2 SINGLE SPATIOTEMPORAL RANDOM FUNCTION MODEL

The single spatiotemporal random function model was chosen to analyse the standardised king prawn catch rate data. With the objective of estimation for a future season and four seasons of data available (2001-2004), it was decided to use the first three seasons (2001-2003) to provide a model that could be used for prediction in the fourth season (2004). In reality, spatiotemporal behaviour of the following season is unknown and the use of a model for a future season based on the behaviour of preceding seasons is justified by the persistence of behaviour over the preceding seasons and a reasonable assumption that those patterns should continue into the next season. Although not undertaken for this study, the environmental (climate and oceanographic) data in 2004 could be examined to confirm that they were consistent with those of 2001-2003. Whilst the 2004 season was to be forecast, the analysis of its spatiotemporal behaviour is also presented along with the 2001-2003 seasons to enable comparison.

Temporal trend analysis of the 2001-2003 data was used to predict the temporal trend model, which was then used to transform the 2001-2003 *CPUE* data to *CPUEs* data. Spatiotemporal variography of the 2001-2003 *CPUEs* data was used to construct a spatiotemporal semivariogram model that was used in a kriging procedure (ST-OK) to estimate the *CPUEs* variable in the future season 2004. The estimated *CPUEs* values were transformed to *CPUE* values using the temporal trend model defined previously.

As the catch rates exhibited a trend over space and time, a variant of kriging known as Universal Kriging (UK), also referred to as kriging with a trend (KT), was considered that allows the modelling of a non-constant trend within the kriging process. This kriging variant may be used when it is inappropriate to consider the local mean as constant even within small neighbourhoods. However, implementation of UK using a global search neighbourhood was unrealistic for the prawn catch rates due to the large number of data. Consideration of the UK trend as a smoothly varying function of the coordinates within the local neighbourhood was also problematic as the prediction surface of the trend is not continuous. Ordinary kriging with a local neighbourhood was seen as the preferred alternative to universal kriging with a global neighbourhood as the mean was allowed to vary smoothly at the larger scale, and was estimated locally at each spatiotemporal location from the data lying in the neighbourhood.

As there was a 4 to 5 month separation between consecutive seasons and a temporal range of continuity of less than one month, estimation of catch rates for a particular month of the 2004 season were produced using data up to but not including the month being estimated.

The 2004 catch rate values used as input data were first transformed to *CPUEs* values using the temporal trend model derived from the 2001-2003 data, and then used in a ST-OK estimation procedure with the spatiotemporal variogram derived for the 2001-2003 season. Estimation was conducted for all lunar months of 2004 except the first lunar month of the fishing season, i.e. April through October.

In order to investigate the uncertainty of the catch rate values, a simulation study was also conducted to simulate catch rate values for the 2004 season using the 2001-2003 data to construct the model. The temporal trend model defined previously was used to transform the data. As the simulation procedure used (ST-SGS) required Gaussian data, the spatiotemporal semivariogram model used in the simulation procedure was constructed from the normal scores (*CPUEsN*) of the standardised *CPUEs* data. As for the estimation procedure, a ST-SGS simulation procedure was conducted for all lunar months of 2004 except the first month, as the first month's data of 2004 (transformed to *CPUEsN* variables) was required as input.

4.3.3 SPATIOTEMPORAL VARIOGRAPHY

Spatiotemporal semivariograms of the standardised king prawn catch rate of the individual seasons 2001 to 2004 were first modelled. The semivariogram models showed many similarities across the individual seasons, which supported the modelling of spatiotemporal semivariograms of combined consecutive seasons. It was initially hoped to compare the prediction of CPUE values for season 2004 using the semivariogram model of the individual season 2003 with the prediction attained using the model for the combined season 2001-2003. However, as the temporal trend model and spatiotemporal semivariogram model of the individual season of 2003 showed some inconsistencies with the other seasons, only the model of the combined season 2001-2003 was used as an estimate for the semivariogram of the standardised catch rate for the 2004 season.

The single spatiotemporal random function model used spatiotemporal variograms computed and modelled via an appropriate joint space-time variogram/covariance model. Spatiotemporal semivariograms and their models were investigated for the *CPUEs* variables of both methods of aggregation, by vessel and by grid. Spatiotemporal variography was conducted for the *CPUEsvd* and *CPUEsgd* variables of the three individual seasons 2001 to 2003 along with the three combined seasons 2001-2002, 2001-2003 and 2002-2003. It is noted that although the 2004 season was to be used as a forecast season, spatiotemporal variography of the 2004 season was still conducted for comparison with the other seasons. Similarly, the combined

(consecutive) seasons 2001-2004, 2002-2004 and 2003-2004 involving the forecast season 2004 were also investigated.

4.3.4 CROSS VALIDATION

Cross validation of the semivariogram model was undertaken to assess how well the semivariogram model of the 2001-2003 combined season captured the spatiotemporal variation of the *CPUEs* data. The cross validation procedure was outlined previously and involves estimating at a spatiotemporal sample point after excluding the sample value at that spatiotemporal location. As the model was to be used for estimating into the 2004 season, cross-validation was conducted for all lunar weeks of the 2004 season to be used as input data for space-time estimation of the subsequent month.

4.3.5 SPATIOTEMPORAL ESTIMATION AND SIMULATION

The selected semivariogram model (combined season 2001 to 2003) was considered in an ST-OK procedure and ST-SGS procedure to predict 2004 values. Spatiotemporal estimation and simulation of the *i*th month uses estimation year data up to, but not including, the *i*th month and the relevant semivariogram model. Estimation (Chapter 8) and simulation (Chapter 9) are conducted for all months except the first month of the fishing season, ie. April through October. A re-estimation process was conducted for the May lunar month of 2004 using survey data to provide relevant input data in order to estimate into a previously unopened fishing area (ENA). A flow chart showing the detrending, variography, estimation and simulation processes (Figure 4.2) denotes the modelling framework presented in this study.

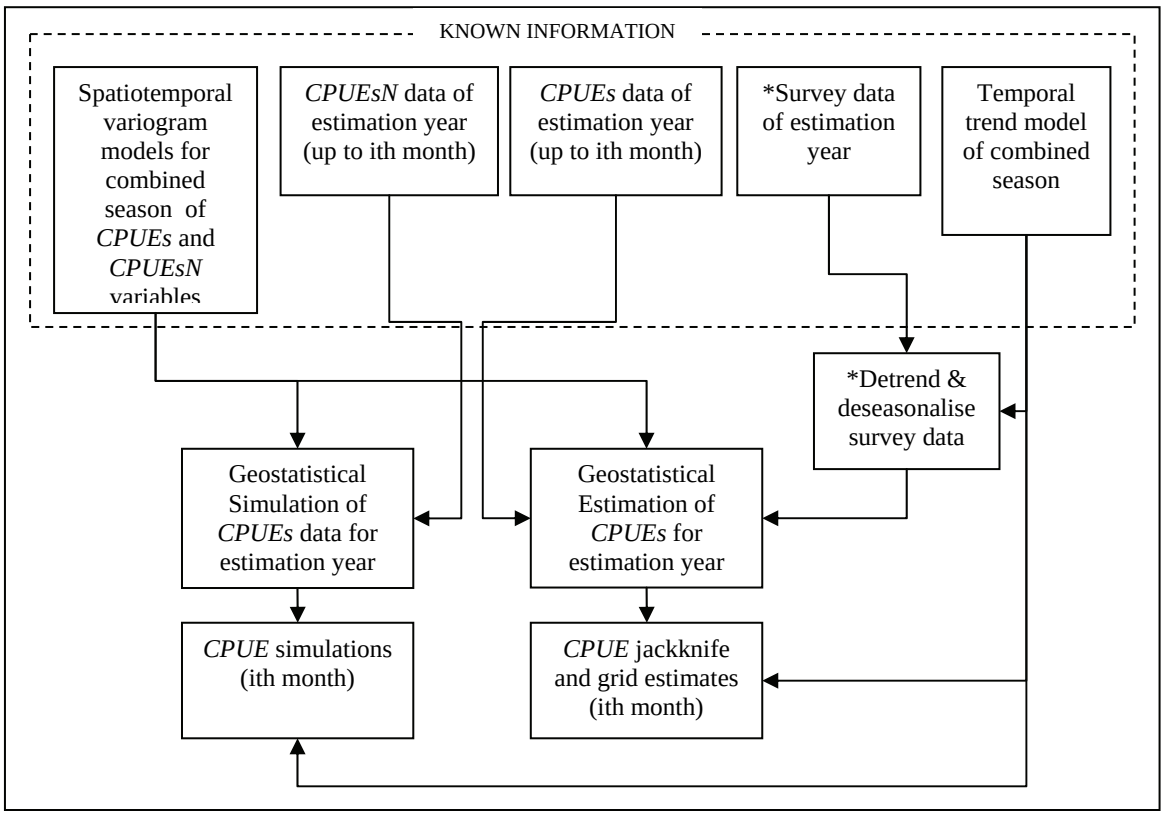


Figure 4.2. Flow chart for spatiotemporal estimation and simulation of *i*th month. *Survey data only used as additional data to estimate for lunar month of May 2004.

4.4 SOFTWARE

MATLAB and EXCEL were the main software packages used in this study in order to conduct exploratory data analysis and to produce plots for the thesis. The geostatistical algorithms used in this research were implemented using the GSLIB package of geostatistical routines. Various routines were used from the GSLIB package including *gamv* and *nscore*. Modified *gamv* and *kt3d* routines (De Cesare et al., 2002) were used to compute space-time semivariograms and to conduct space-time estimation and a modified version of *sgsim* was developed in order to implement spatiotemporal Gaussian simulation. MATLAB and EXCEL were used to conduct post-processing of outputs from the GSLIB geostatistical routines.

PART II.

GLOBAL TREND MODELLING AND SPATIOTEMPORAL CONTINUITY

CHAPTER 5

CATCH RATE STANDARDISATION

5.1 INTRODUCTION

Catch rates of aquatic populations, based on analyses of information on catch and effort within a fishery are an important indicator variable for use in a fishery (Walters, 2003). The utility of indices based on catch and effort data is improved by standardising them to remove the impact of factors such as changes over time in the composition (and hence efficiency) of the fleet. By removing as much of the variability due to vessel differences, seasonal and time effects as possible, catch and effort data from different vessels can be compared.

The standardisation procedure is utilised in this study to remove a source of spatiotemporal non-stationarity (non-constant temporal mean and temporally-varying variance). After removing the large-scale temporal variability, catch rates from across the season can be considered relative to each other. The local behaviour in catch rates can then be investigated via the spatiotemporal structure of the residuals. In particular, catch rates will be standardised by lunar month and phase so that they can be used in a geostatistical spatiotemporal model.

A previous study involving the king prawn catch rate data in Shark Bay indicated that the annual time series could be modelled via a multiplicative classical decomposition model (Harman, 2001). The study involved annual series of daily catch rate values from six non-consecutive years from 1972 to 1987 and showed that the king prawn catch rate data had a seasonal factor associated with the lunar phase. The detrended daily data had a 29-day cycle that corresponded to the lunar index cycle. The catch was at a low at the full moon and at a high at the new moon.

Median polish trend modelling is also used to investigate the inclusion of a spatial component, in addition to the temporal component, in the global trend model. This trend modelling process has been used in another geostatistical space-time trend analysis (Bruno, 2004).

As outlined in the methodology, the modelling framework of this study aims to develop a model from the 2001-2003 seasons that can be used to predict catch rates of the 2004 season. Temporal trend models are presented in this chapter for the 2001-2004 seasons, but it must be

noted that the temporal trend used in the predictive model for 2004 will be based only on the 2001-2003 seasons. The temporal trends of 2004 and the associated combined seasons are shown for comparison, whilst in reality when predicting a future season such comparisons would not be possible.

5.2 TEMPORAL TREND ANALYSIS

Weekly averaged king prawn catch rate data (*CPUEw*) of seasons 2001 to 2004 were used to allow comparison of temporal models across the seasons. Temporal trends on a 10 nmil by 10 nmil grid within the study region were also investigated but did not provide any more useful information than trends calculated over the entire study region.

The temporal trends of combined seasons of data were also investigated by combining data for each lunar week over successive seasons. This was possible as the weeks of each season were numbered so they were aligned by moon phase and the approximate alignment of calendar months (Section 3.4.1). The investigation of combined seasons clarified the persistence of temporal trends of individual seasons and enabled the construction of a trend model for a future season based on the combination of the preceding seasons The following six combined sets of consecutive seasons were considered: 2001 to 2002, 2001 to 2003, 2001 to 2004, 2002 to 2003, 2002 to 2004 and 2003 to 2004. The persistence of a trend across a number of seasons justifies its use in prediction of a future season. This meant that a trend model obtained for the 2001-2003 seasons could be used to model the 2004 season.

5.2.1 SPECTRAL ANALYSIS

The weekly king prawn catch rate data of this study were analysed using spectral analysis via the ASTSA program to check if the cyclic behaviour of the king prawn CPUE noted previously (Harman, 2001) is present in these data. Spectral analysis requires complete series whilst the annual series of 2001 to 2004 contain a number of missing values (Figure 3.3). These values were estimated by using the centred moving average value for that week.

The spectra of the amended king prawn CPUE for the individual seasons (Figure 5.1, left) all have a strong peak at a low frequency (0.0625) and another peak at the 0.25 frequency. A frequency of 0.0625 corresponds to a period of 16 while the length of the data set is only 30 in 2003, and 31 in the other seasons. A peak at such a relatively low frequency typically indicates the presence of an important trend component (Chatfield, 2004). The other peak indicates the presence of a cyclical component. By estimating the trend using a 4 point centred

moving average corresponding to the 4 week lunar cycle and assuming a multiplicative trend, the centred moving average is divided into the CPUE data to obtain a detrended data set.

Spectra of the detrended king prawn CPUE (Figure 5.1, right) no longer show a peak at low frequency supporting the presence of a trend estimated by the centred moving average. All individual seasons still display a strong peak at the 0.25 frequency corresponding to a cyclical component with a 4 week period. Therefore, the assumption of a 4 week lunar cycle is justified.

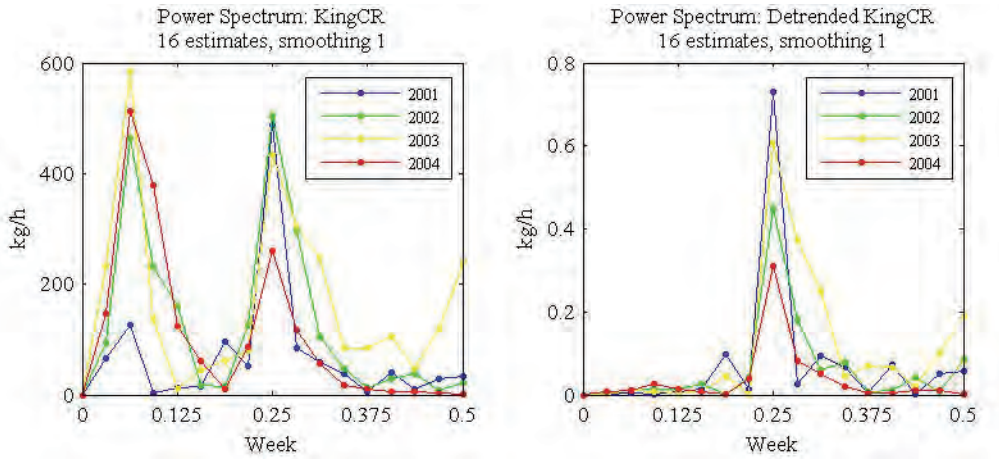


Figure 5.1. Spectrums of original (left) and detrended (right) king prawn CPUE for individual seasons, 2001-2004.

Spectra for the combined seasons (Figure 5.2, left) also display a strong peak at a low frequency (0.0625) indicating the presence of a trend and another peak at the 0.25 frequency corresponding to a cyclical component with a 4 week period. After detrending, (Figure 5.2, right) only a strong peak at the 0.25 frequency is evident and therefore, as for the individual seasons, the assumption of a 4 week lunar cycle is justified.

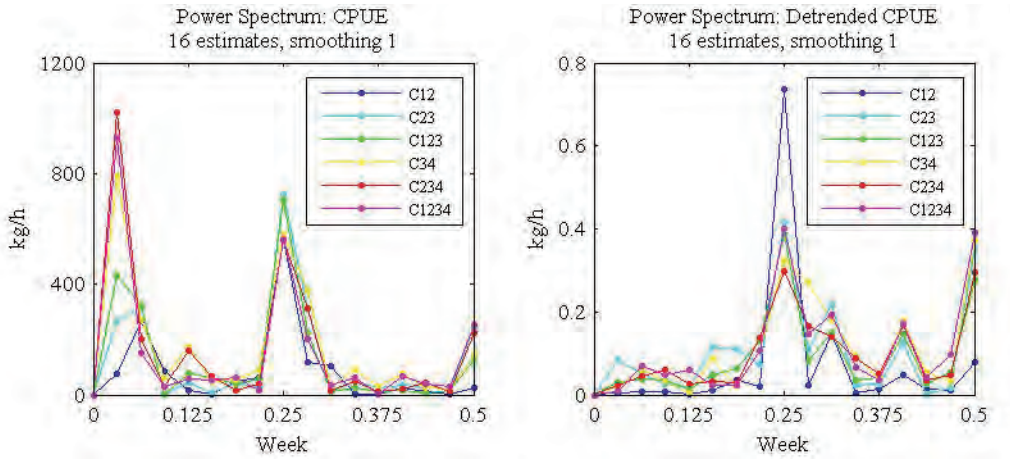


Figure 5.2. Spectrums of original (left) and detrended (right) CPUE for combined seasons, 2001-2004.

5.2.2 CENTRED MOVING AVERAGES

The 4-point centred moving averages of the weekly averaged king prawn CPUE series for the four consecutive seasons 2001-2004 show a peak in the first half of the season and a levelling

off at much lower values by the end of the season (Figure 5.3). Centred moving averages of the weekly averaged king prawn CPUE series for the six combined seasons show a similar pattern (Figure 5.4). All four individual seasons and six combined seasons show seasonal fluctuations about their centred moving average with differences in the variability of their values. A multiplicative model appears reasonable as the seasonal pattern is proportional to the data.

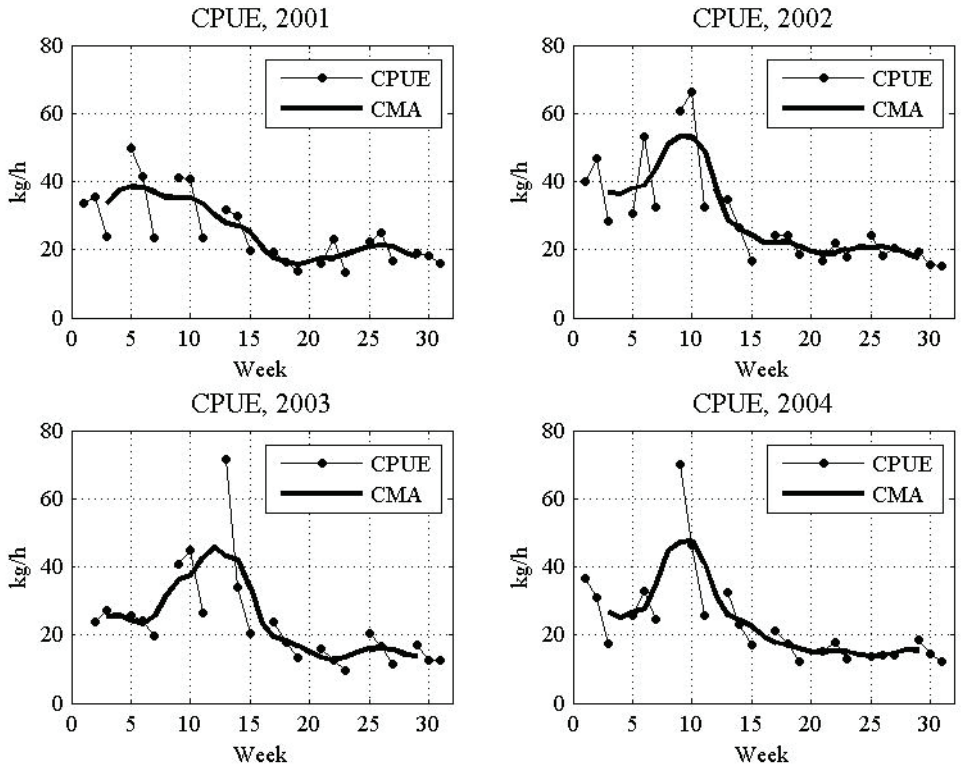


Figure 5.3. Weekly averaged king prawn CPUE with 4 point centred moving average, 2001-2004.

A comparative chart of the centred moving averages for each year (Figure 5.5, left) shows that 2002 and 2004 have a similar peak around week 9, in line with the opening of the Carnarvon-Peron line. As the prawns are typically more vulnerable for a short time after re-opening an area that has been left unfished for a period, it is expected that there will be a short period of greater CPUE. However, the main reason for the increase in CPUE is the fact that new recruits have entered the fishery. The line is opened when the newly-recruiting prawns are of a sufficient size, timing of which varies slightly with fishing season. In season 2003 there is a peak in week 13, again in line with the opening of the Carnarvon-Peron line, a few weeks after the peak of 2002 and 2004. Season 2001 does not have a significant peak. For seasons 2001 and 2002 the starting values are higher than in the other seasons and these two years show generally higher values than the other years.

The combined seasons show a similar shape to the individual seasons with a peak of lower magnitude in the first half of the season (Figure 5.4). The starting values are higher for those combined seasons involving 2001 and 2002 and lower for those involving 2003 and 2004 (Figure 5.5, right).

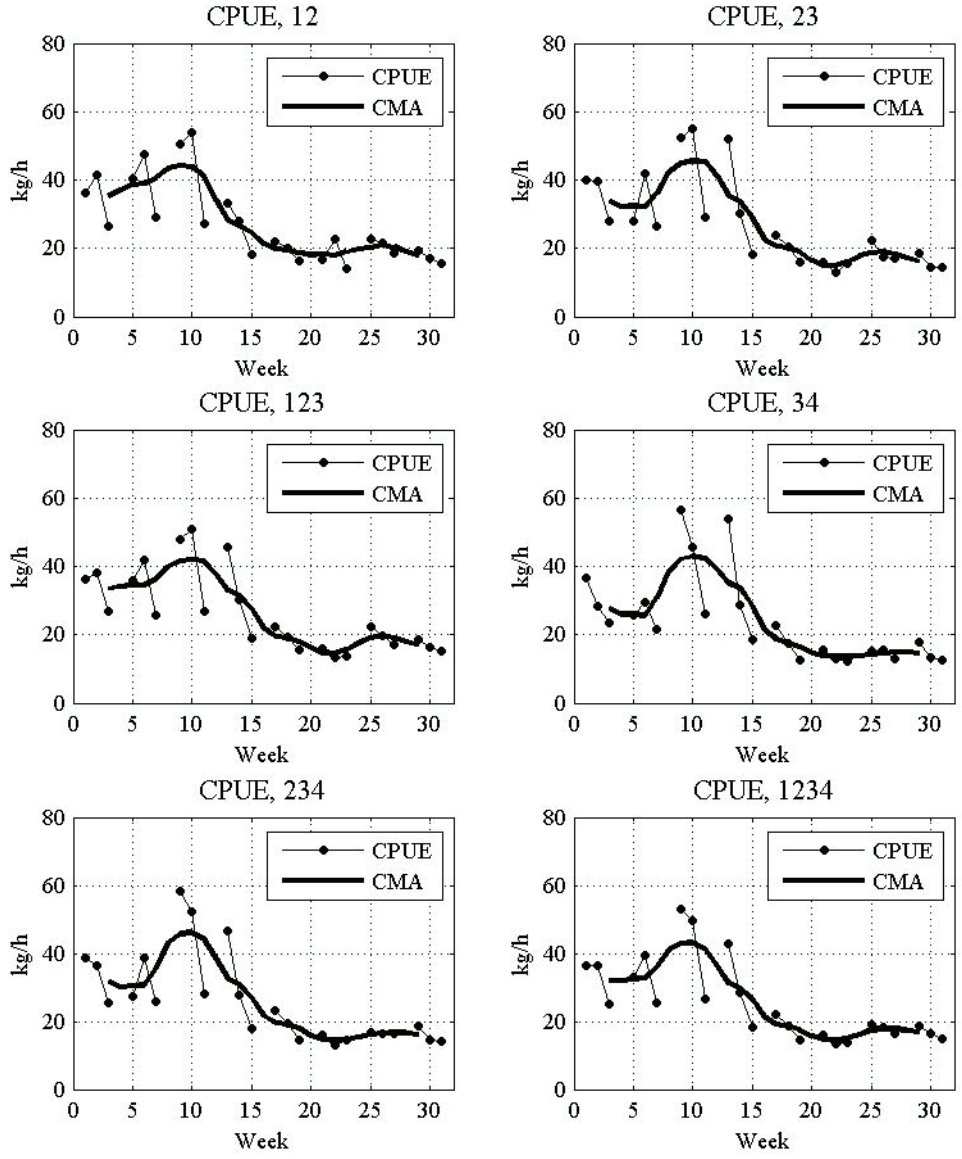


Figure 5.4. Centred moving average, combined seasons, 2001-2004.

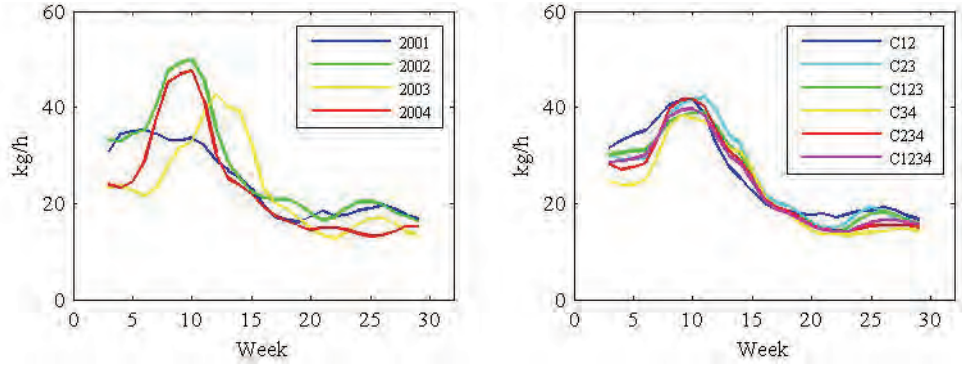


Figure 5.5. King prawn CPUE 4 point centred moving averages, individual (left) and combined (right) seasons, 2001-2004.

5.2.3 CLASSICAL DECOMPOSITION MODEL

We assume a multiplicative classical decomposition model for each season as the seasonal factor was proportional to the trend in all seasons (Figure 5.3). The model for each year consists of a trend component, a single seasonal component of length 4 weeks and a (multiplicative) residual component.

$$Y(t) = T(t) \cdot S(t) \cdot R(t) \tag{5.1}$$

No cyclical component is considered as the series is annual.

5.2.3.1 SEASONAL INDEX

The four "seasons" correspond to the four lunar phases: New Moon (N), First Quarter (Q), Full Moon (F) and Last Quarter (L). By assigning an index number to each lunar phase (Table 5.1) we can assign a season number $i = 1,2,3,4$ to each time instant using

$$i(t) = t - 4\left\lfloor \frac{t-1}{4} \right\rfloor \tag{5.2}$$

This assigns the numbers 1 to 4 (Table 5.1) to each of the weeks associated with a lunar phase. As there are four seasons evident in each annual time series, the 4 point centred moving average calculated previously is used as an initial estimate of the trend component for each season. This centred moving average is divided into the data to obtain the seasonal irregular component. The average seasonal index is then calculated and normalised for each of the four seasons $i = 1,2,3,4$ in each season. Note that there are no data for those weeks corresponding to the full moon weeks. The data from these weeks have been absorbed into the adjacent weeks due to the full moon closures.

Table 5.1. Seasonal index numbers for lunar phases.

Phase	L	N	Q	F
Index	1	2	3	4

The classical decomposition seasonal indices for the king prawn CPUE of each season show strong similarities in the factors between years (Figure 5.6, left). All seasons showed the lowest annual factors for the first quarter moon week. Seasons 2001 and 2002 had the largest annual factor for the new moon week and the second largest factor for the last quarter week. For 2003 and 2004 these positions were reversed. The magnitudes of the annual factors for each lunar phase are quite similar across the seasons. The magnitude of the last quarter moon week is

approximately 0.75 while those of the last quarter moon and new moon weeks vary between 1.00 and 1.25.

All combined seasons showed the lowest annual factors for the first quarter moon week (Figure 5.6, right). With the exception of the combined season 2001-2002, all combined seasons showed the largest annual factor for the last quarter moon week and the second largest factor for the new moon week. This may be due to the influence of the relatively high/low values for the last quarter/new moon weeks respectively of season 2003, as season 2003 is present in all combined seasons except 2001-2002. The magnitudes of the annual factors for each lunar phase are quite similar across the seasons. Similar to the case for the individual seasons, the magnitude of the last quarter moon week is approximately 0.75 while those of the last quarter moon and new moon weeks vary between 1.00 and 1.25 (Table 5.2).

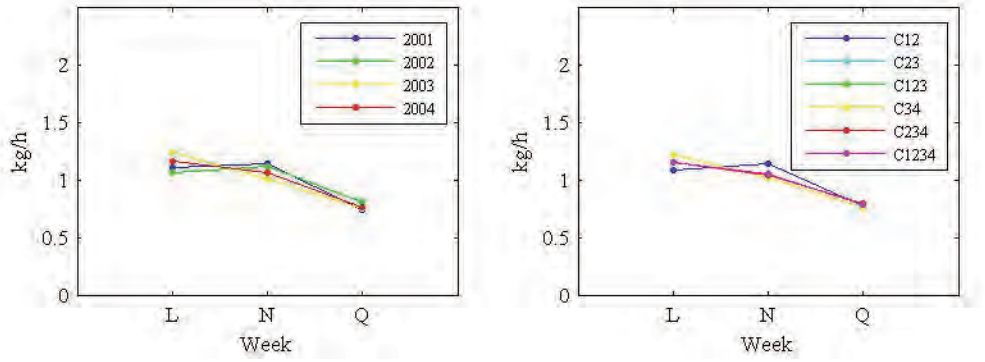


Figure 5.6. King prawn CPUE seasonal indices, individual (left) and combined (right) seasons, 2001-2004.

Table 5.2. Seasonal index values for individual and combined seasons.

Season	Last Quarter (L)	New Moon (N)	First Quarter (Q)
2001	1.114	1.139	0.747
2002	1.063	1.127	0.810
2003	1.244	1.007	0.750
2004	1.172	1.061	0.767
C12	1.087	1.141	0.772
C23	1.158	1.042	0.802
C123	1.158	1.058	0.784
C34	1.217	1.018	0.766
C234	1.157	1.040	0.804
C1234	1.159	1.052	0.789

5.2.3.2 DESEASONALISED TREND

The deseasonalised data for the individual and combined seasons of 2001 to 2004 were obtained by dividing the original data by the relevant seasonal factor, which was then fitted with a polynomial trend line (Figure 5.7) using the method of least squares. For all individual and combined seasons a cubic function was considered to be appropriate for modelling the trend

(Table 5.3). It was noted that the cubic function indicated an increase at the end of the season. This may be explained by temperature beginning to increase at this time which would make prawns more active and therefore more vulnerable. In addition, there would also be a tendency for prawns to aggregate in spawning areas, also making them more vulnerable. The deseasonalised trends for seasons 2001 to 2004 show similar shapes for all four seasons (Figure 5.9, left). In general there is a decrease in CPUE towards the second half of the year. The deseasonalised trends compare favourably to the centred moving averages of the data (Figure 5.5). Peaks of the deseasonalised trends were less pronounced for seasons 2002 to 2004 and the centred moving averages were much flatter in shape for the second half of the season.

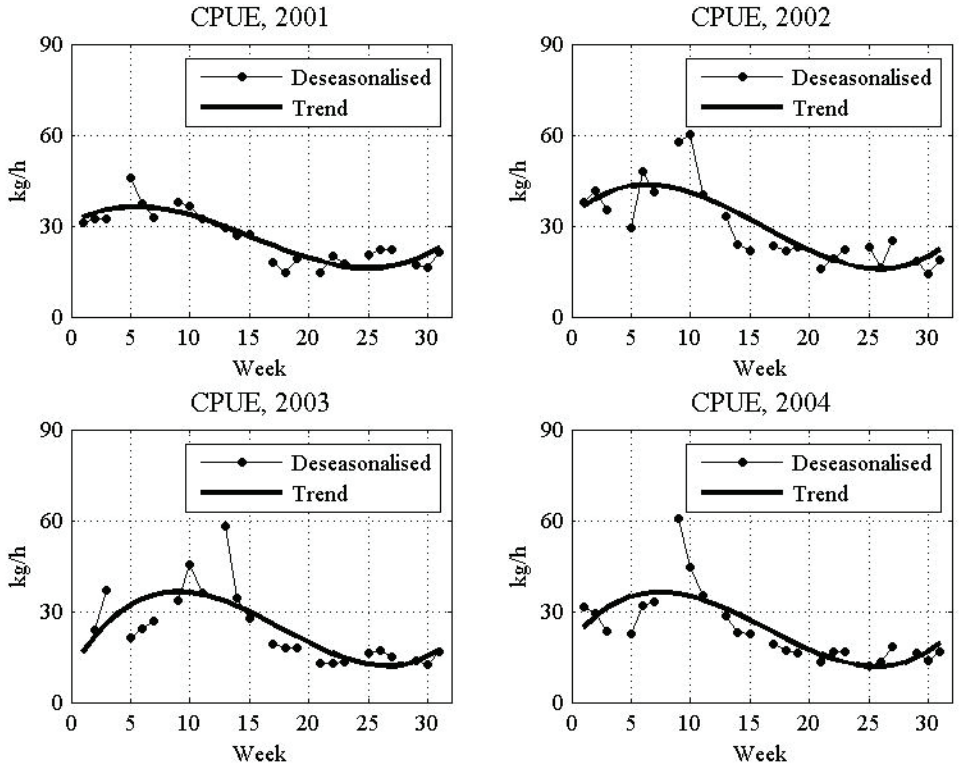


Figure 5.7. Deseasonalised king prawn CPUE with fitted polynomial trend, 2001-2004.

Table 5.3. Fitted cubic deseasonalised trend models for individual and combined seasons.

Season	Deseasonalised Trend	R^2
2001	$0.00521\,x^3-0.236\,x^2+2.072\,x+30.309$	0.776
2002	$0.00699\,x^3-0.337\,x^2+3.360\,x+33.231$	0.649
2003	$0.00887\,x^3-0.473\,x^2+6.366\,x+10.352$	0.563
2004	$0.00783\,x^3-0.388\,x^2+4.450\,x+20.460$	0.800
C12	$0.00636\,x^3-0.298\,x^2+2.844\,x+31.618$	0.792
C23	$0.00700\,x^3-0.345\,x^2+3.693\,x+28.266$	0.689
C123	$0.00702\,x^3-0.337\,x^2+3.519\,x+27.781$	0.781
C34	$0.00767\,x^3-0.390\,x^2+4.657\,x+19.584$	0.681
C234	$0.00760\,x^3-0.374\,x^2+4.091\,x+25.507$	0.703
C1234	$0.00737\,x^3-0.356\,x^2+3.796\,x+25.974$	0.773

The shapes of the fitted deseasonalised trends for the combined seasons (Figure 5.8) are similar to those of the individual seasons (Figure 5.7). The variability of the catch rate data from the individual season trend models is greatest at the beginning of the season and decreases over the season. The trends of the combined seasons show much less variability than the individual seasons (Figure 5.9). The combined seasons involving the latter seasons (2003 and 2004) display a lower trend across the year than those involving the earlier seasons (2001 and 2002).

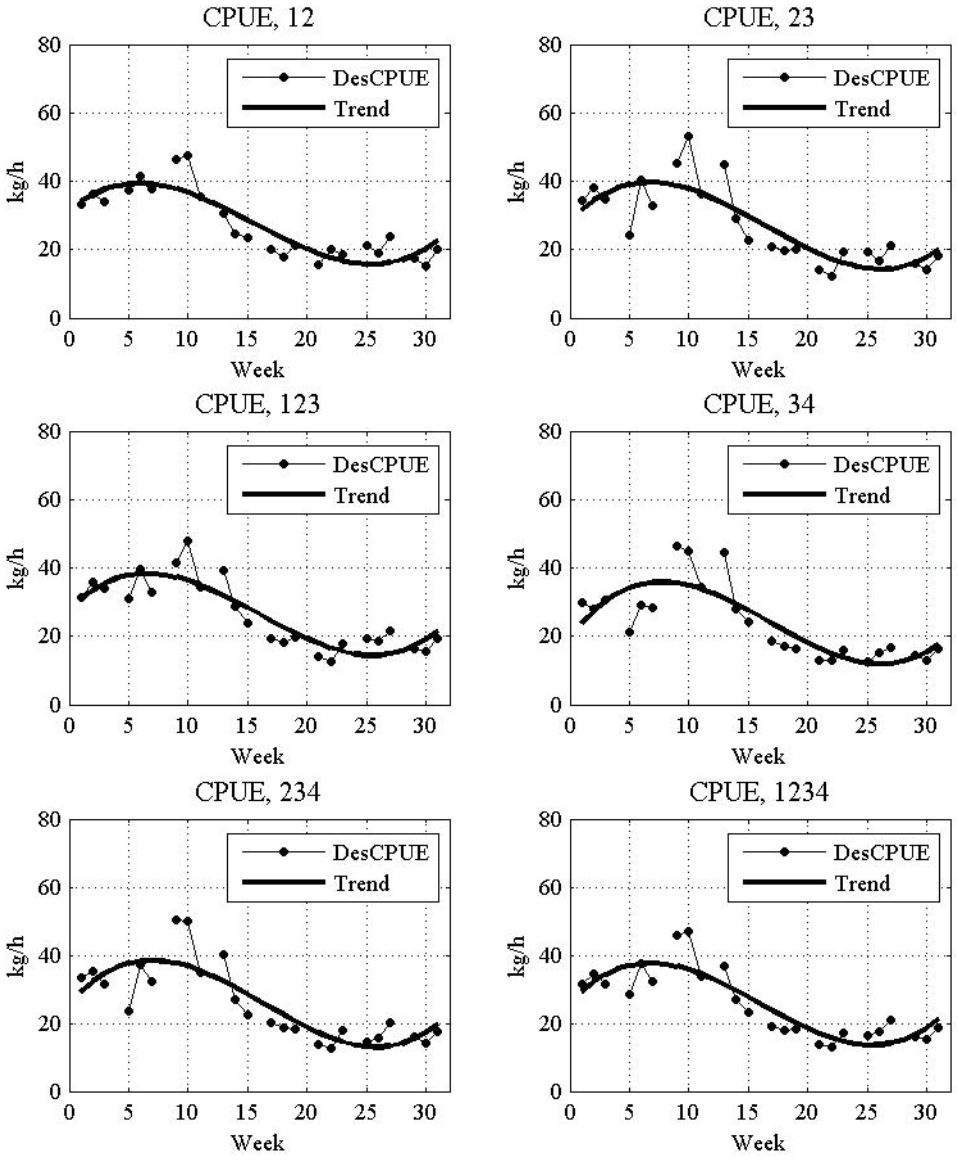


Figure 5.8. Deseasonalised king prawn CPUE with fitted polynomial trend, 2001-2004.

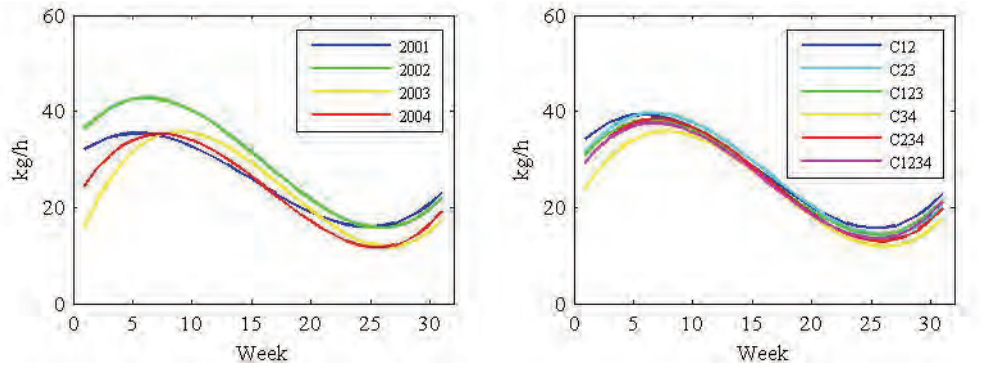


Figure 5.9. King prawn CPUE deseasonalised trend models, 2001-2004.

5.2.3.3 MODEL FIT AND ERROR ANALYSIS

Classical decomposition models are obtained by multiplying the deseasonalised trend by the seasonal factor for the relevant individual and combined season. The classical decomposition models for the individual seasons 2001 to 2004 (Figure 5.10) show that the largest errors correspond to the main peaks observed in the original data (Figure 5.3). Further inspection reveals that each of these peaks corresponds to the opening of either the Carnarvon-Peron Line or the Extended Nursery Area (Table 5.4). Both these areas protect emerging young prawns from being caught too early and therefore explain the peaks evident as the fishermen target these areas as they are opened.

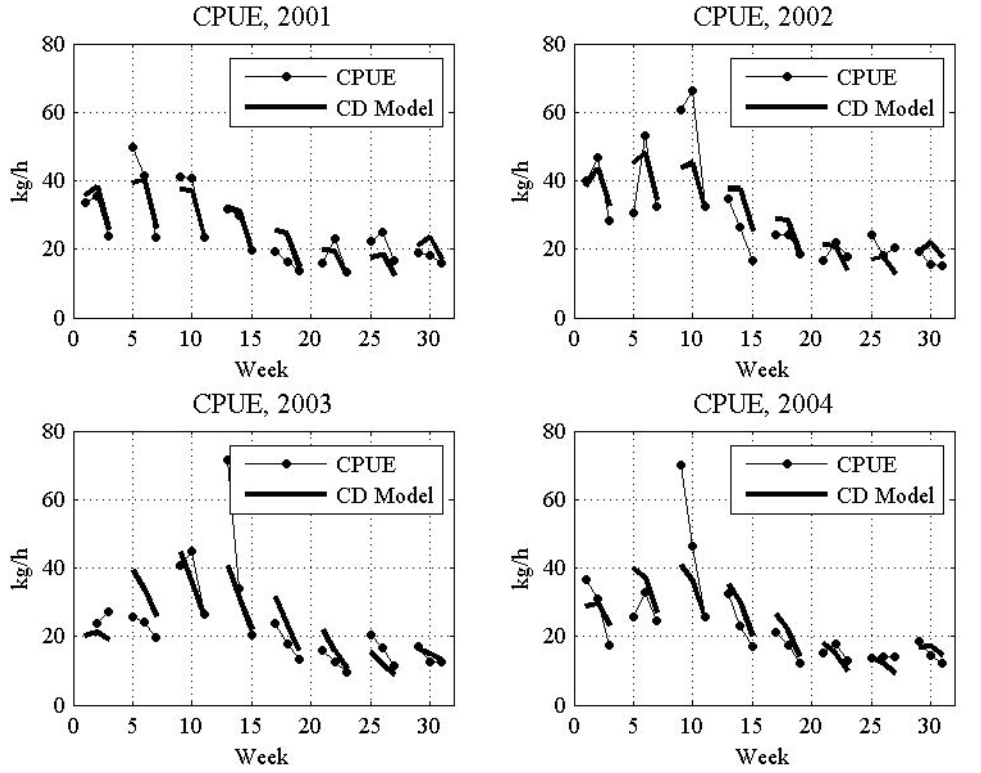


Figure 5.10. King prawn CPUE classical decomposition model, 2001-2004.

Table 5.4. Peaks of weekly averaged king prawn CPUE and corresponding line openings.

Season	Main peak of season	Line opening
2001	Week 5	CP
2002	Week 9	ENA
2003	Week 13	ENA
2004	Week 9	CP(lower) & ENA

Classical decomposition models for the combined seasons (Figure 5.11) show similarity with the individual seasons. The largest errors coincide with the peaks in weeks 9 and 13 corresponding to the opening of the Carnarvon-Peron Line and ENA evident in the individual seasons. As expected, the models for the combined seasons show less variability between seasons than the individual seasons (Figure 5.12).

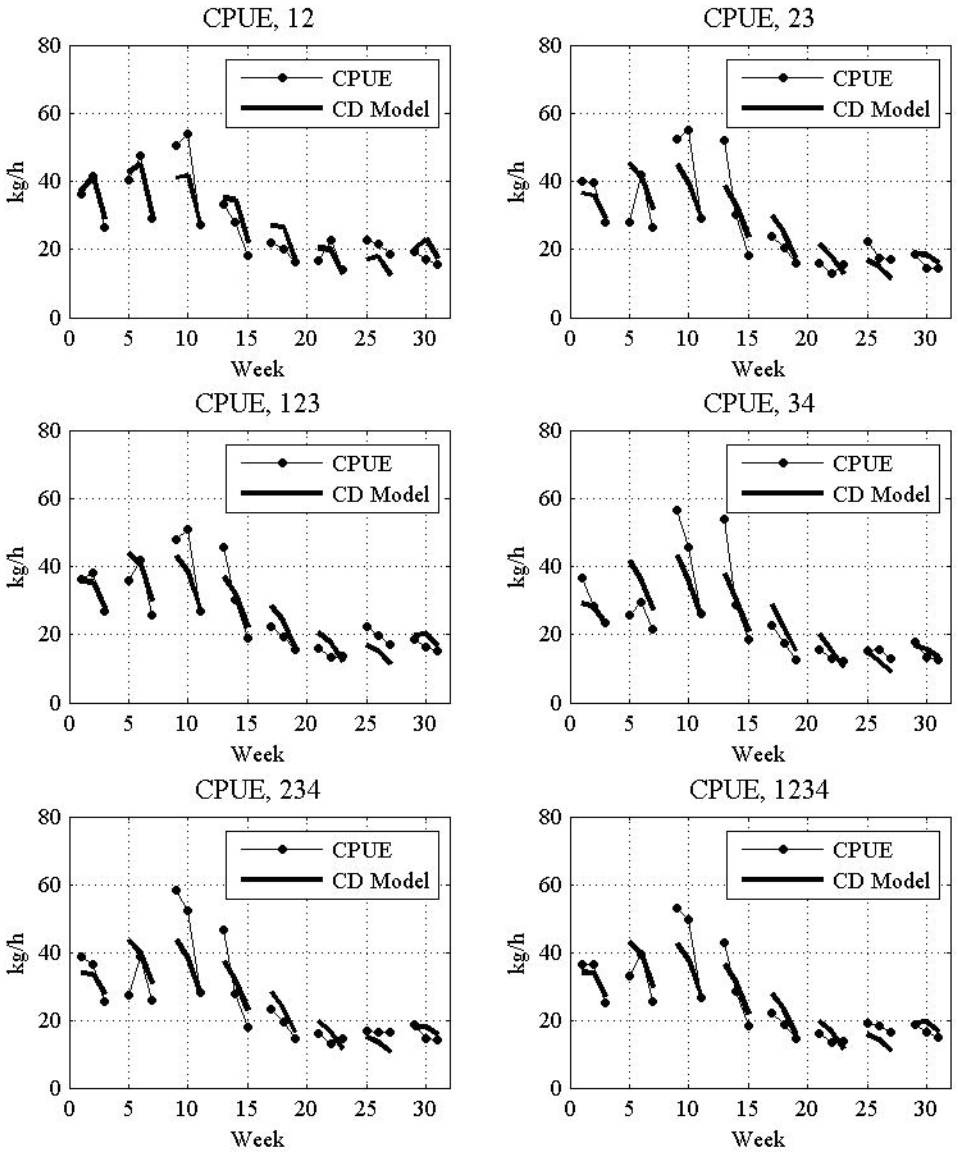


Figure 5.11. King prawn CPUE classical decomposition model, combined seasons, 2001-2004.

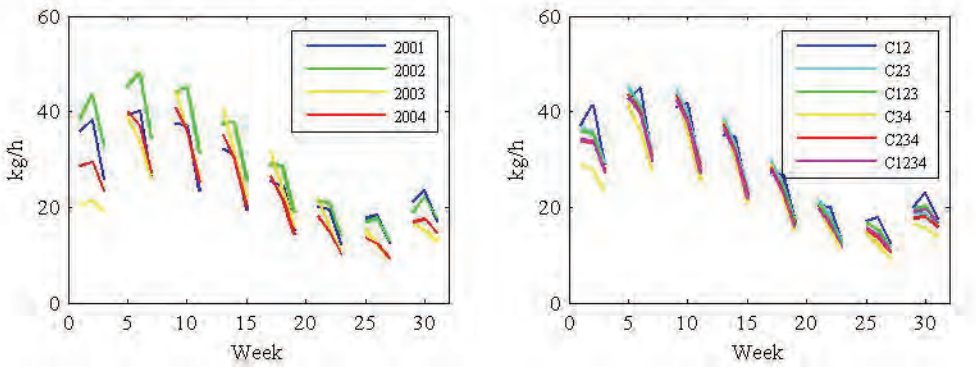


Figure 5.12. King prawn CPUE classical decomposition model, combined seasons, 2001-2004.

Errors for the classical decomposition models (Figure 5.13) show that the 2001 season has the smallest mean error (0.02) and smallest mean percentage error (2.57%) in absolute terms. The 2004 season has the largest mean error (0.18) whilst the model for the 2003 season has the largest mean percentage error (6.10%). The percentage errors are negative for all five seasons reflecting the underestimation of the CPUE by the models. The models for seasons 2002 to 2004 have similar mean absolute deviations varying from 5.11 to 5.70 whilst that of 2001 is significantly lower (3.24). The coefficient of determination is smallest for season 2003 (0.61) and greatest for season 2001 (0.83). The values indicate strong correlation between the model and the CPUE for all seasons.

The mean square errors, mean absolute deviations, mean percentage errors and mean absolute percentage errors of the combined seasons are smaller than those of the individual seasons, with the exception of the 2001 season. The R-squared values of the combined seasons are higher than those of the 2002, 2003 and 2004 individual seasons. Only one of the combined seasons (2001-2002) shows a higher R-squared value than the 2001 season.

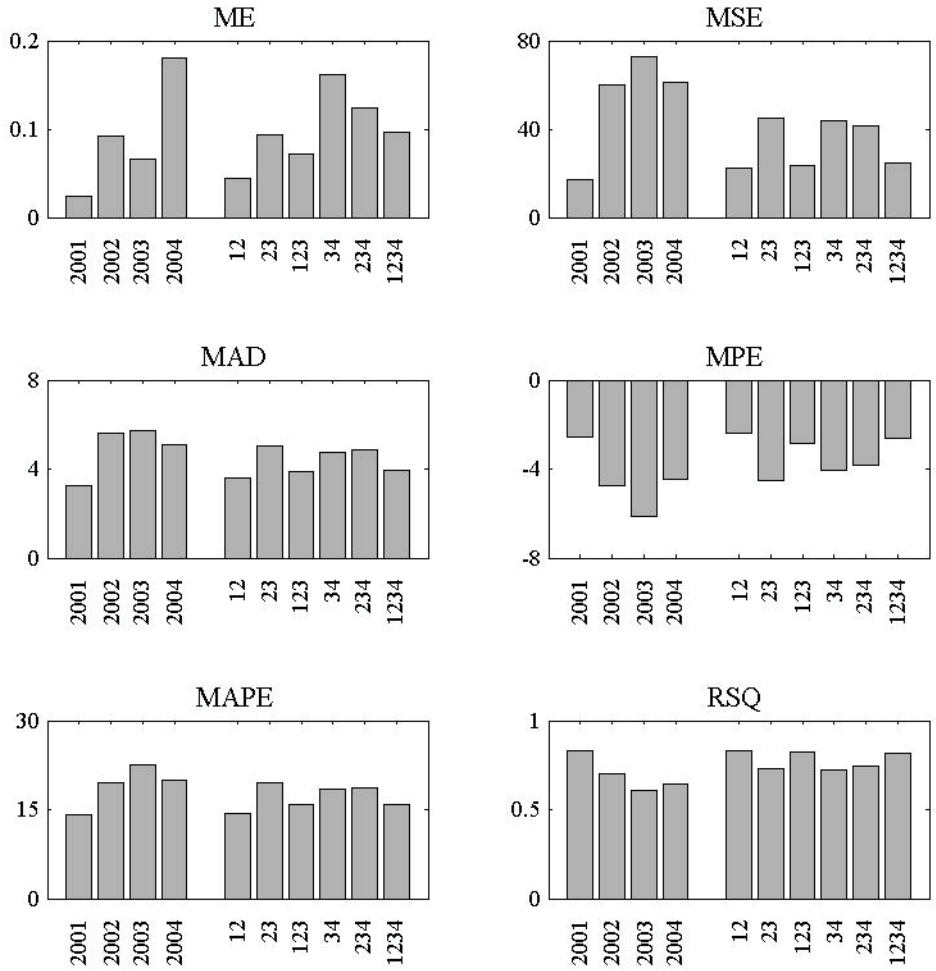


Figure 5.13. Measures of error for the classical decomposition models, 2001-2004.

5.2.4 COMPARISON OF INDIVIDUAL AND COMBINED SEASONS

Fitted deseasonalised trend models for the individual seasons of 2001 and 2004 are very similar to that of the combined seasons involving the relevant individual season (Figure 5.14). In particular, combined season trends for the 2002 season underestimate the individual trend for the first half of the season and those of season 2003 overestimate the individual trend at the beginning of the season. Similar behaviour can be seen for the classical decomposition models of the individual seasons and associated combined seasons (Figure 5.15). The models for the individual seasons 2001 and 2004 are well represented by the models for the combined seasons involving those seasons. As for the deseasonalised trend models, the models of the combined seasons involving season 2002 underestimate the individual model for the first two lunar months of the season. The combined season models involving 2003 overestimate the individual model for the first two lunar months of the season.

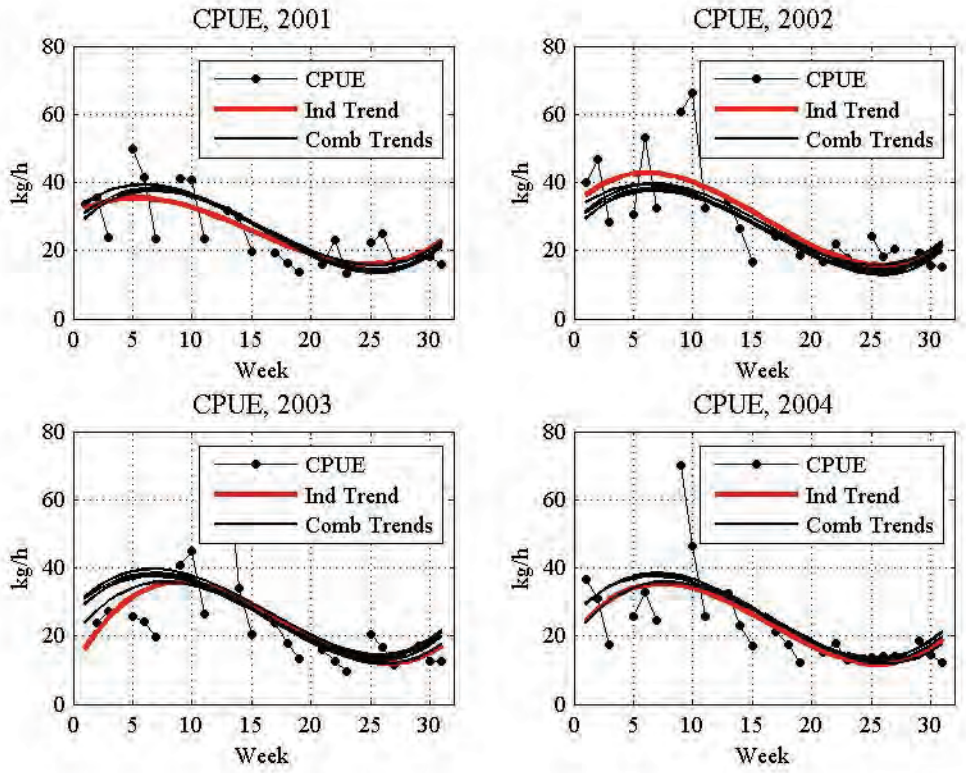


Figure 5.14. Deseasonalised trend models for individual and combined seasons, 2001-2004.

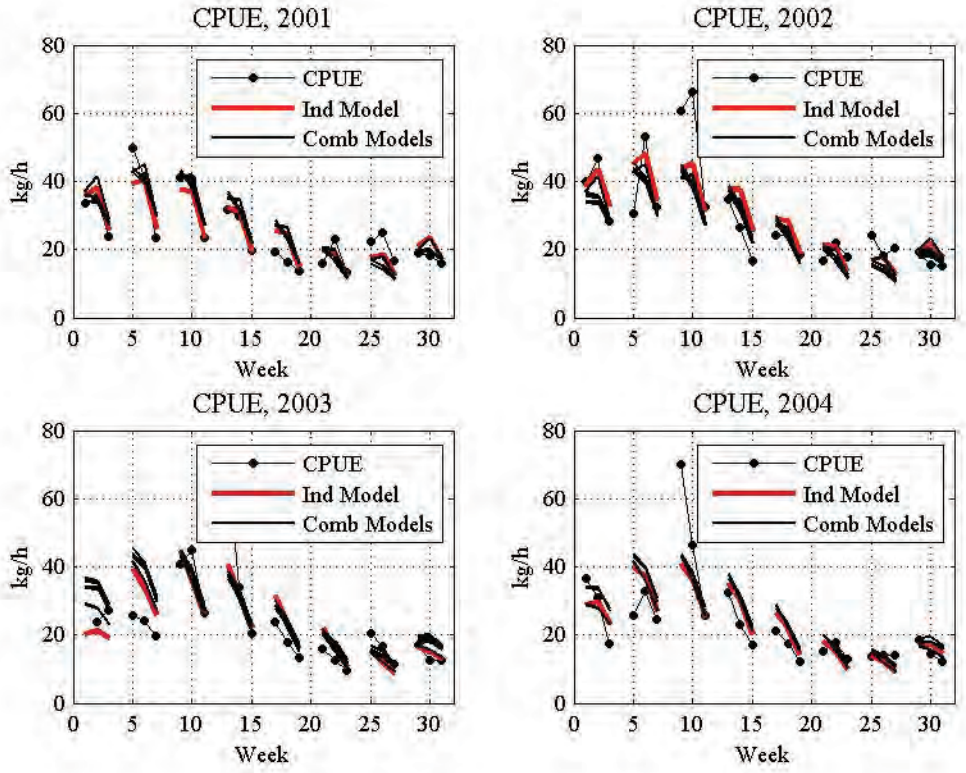


Figure 5.15. Classical decomposition models for individual and combined seasons, 2001-2004.

5.3 SPATIOTEMPORAL TREND ANALYSIS BY MEDIAN POLISH

A median polish spatiotemporal trend model was investigated to consider a spatial component as well as a temporal component in the trend model. The median polish trend modelling procedure can be used for spatiotemporal trend models involving either additive or multiplicative spatial and temporal components along with a main effect. The multiplicative model was chosen over the additive model after an analysis of the errors from both models. This decision was also supported by the multiplicative trend model used in the classical decomposition model. Using information from the classical decomposition temporal trend modelling process, it was decided to include two temporal components: lunar month and lunar phase. There were 8 lunar months with 3 lunar phases in each (no full moon weeks). Two spatial components were considered incorporating a longitude and a latitude effect. Three grid sizes were considered for the spatial grid, 1 by 1 nmil, 5 by 5 nmil and 10 by 10 nmil by aggregating the CPUE shot data by the appropriate grid size. Results indicated that the 5 by 5 nmil grid gave the model of best fit without oversmoothing.

5.3.1 MULTIPLICATIVE MEDIAN POLISH MODEL

A median polish trend model (Figure 5.16) considering only the month and phase temporal effects was first considered to compare directly with the classical decomposition trend model. The main effect is higher in seasons 2001 and 2002 and lower in seasons 2003 and 2004. The month effect showed a peak in the first half of the season and a levelling off at a much lower level by the end of the season. This was similar to the shape of the centred moving averages and deseasonalised trend model seen previously in the classical decomposition model. The peak occurred in the third month for all seasons, and was highest for season 2004. The phase effect was highest for the last quarter moon phase for all seasons except 2002 when the maximum occurs during the new moon phase. The index for the last quarter moon phase was lowest for all season.

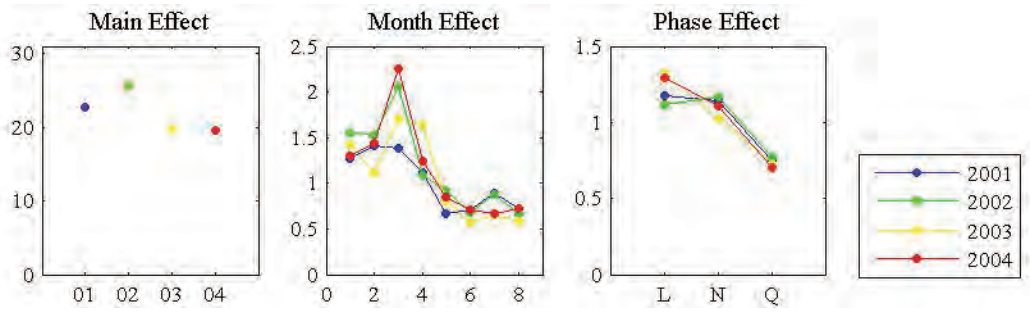


Figure 5.16. Multiplicative median polish spatiotemporal trend effects for individual seasons, 2001-2004.

A comparison of the temporal trend obtained for the median polish model with the classical decomposition model shows that the models perform quite similarly across all four individual seasons (Figure 5.17). There are some notable differences in behaviour of the two models in a few weeks of all seasons such as the third month of season 2002 where the median polish model has higher values than the classical decomposition model. The obvious similarities in the two models confirm the suitability of a multiplicative temporal trend of the catch rates.

The multiplicative median polish model was then extended for the individual seasons to include also the spatial effects (longitude and latitude) with spatial grid size 5 by 5 nautical miles (Figure 5.18). Once again, many similarities were evident across the seasons. The main effect, or mean level of the model, was around 20 for all seasons. Whilst the month effect still levelled off at a lower level for the second half of the season, the peak seen previously in the third month was replaced by a continuously decreasing effect from the start of the season.

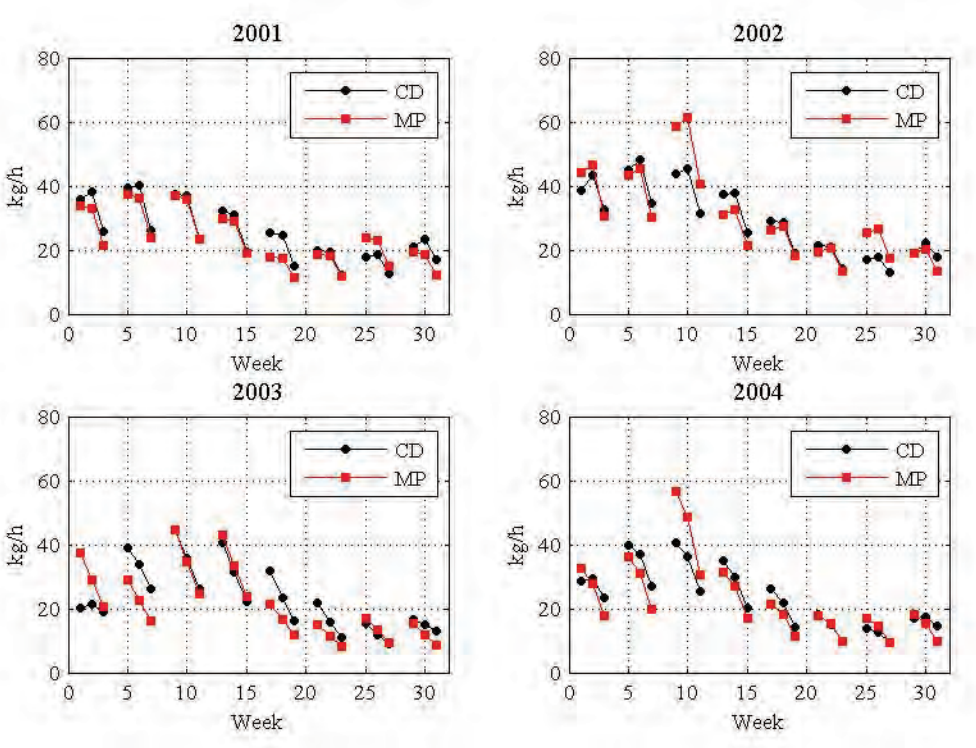


Figure 5.17. Multiplicative median polish spatiotemporal trend effects for individual seasons, 2001-2004.

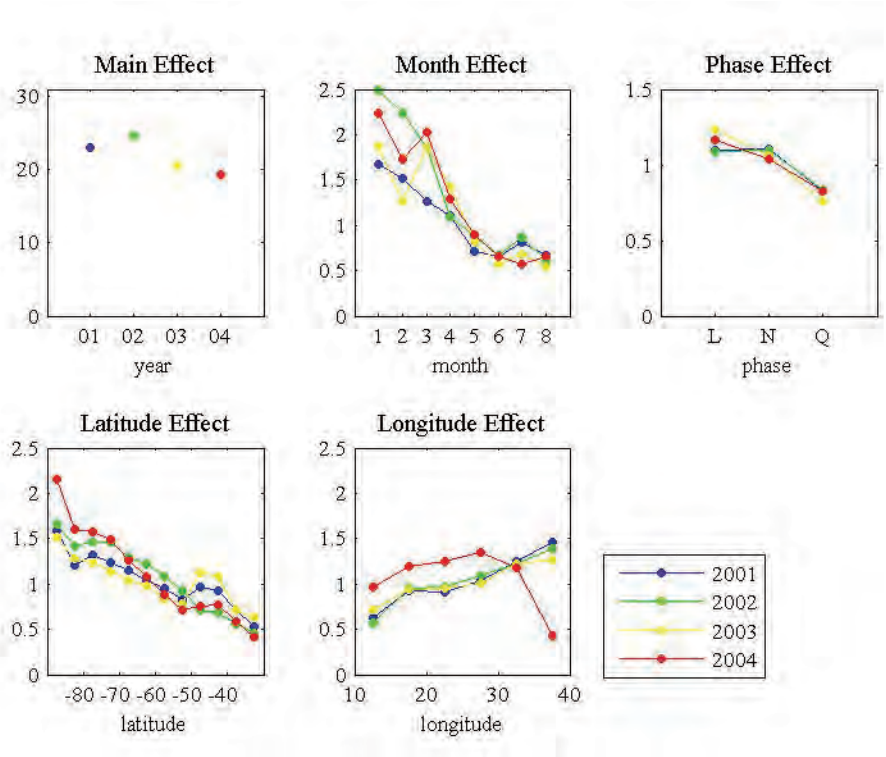


Figure 5.18. Multiplicative median polish spatiotemporal trend effects for individual seasons, 2001-2004.

All seasons showed the lowest annual factors for the first quarter moon week. Seasons 2001 and 2002 had the largest annual factor for the new moon week and the second largest factor for the last quarter week. For 2003 and 2004 these positions were reversed. The magnitudes of the annual factors for each lunar phase are quite similar across the seasons and similar to those seen in the classical decomposition model. The magnitude of the last quarter moon week is approximately 0.8 indicating that the catch rates of these weeks are lower than average. The catch rates of the last quarter moon and new moon weeks are above average with their effects varying between 1.0 and 1.2.

Table 5.5. Seasonal index values for individual and combined seasons.

	Last Quarter (L)		New Moon (N)		First Quarter (Q)	
Season	CD Model	MP Model	CD Model	MP Model	CD Model	MP Model
2001	1.114	1.094	1.139	1.106	0.747	0.826
2002	1.063	1.089	1.127	1.101	0.810	0.834
2003	1.244	1.231	1.006	1.065	0.750	0.762
2004	1.172	1.167	1.061	1.039	0.767	0.825
C12	1.087	1.088	1.141	1.098	0.771	0.837
C23	1.157	1.134	1.042	1.103	0.802	0.800
C123	1.158	1.108	1.058	1.104	0.784	0.817
C34	1.217	1.183	1.018	1.068	0.766	0.791
C234	1.157	1.148	1.040	1.084	0.803	0.804
C1234	1.159	1.121	1.052	1.095	0.789	0.815

The longitude effect showed a steady increase with increasing longitude for seasons 2001 to 2003. For 2004, the longitude effect showed a rapid decrease on the eastern side of the region. The latitude effect showed a steady decrease with increasing latitude for all individual seasons. The combined spatial effect (including latitude and longitude effects) for all four individual seasons show the spatial trend is generally lower in the north-western corner and increases to the south-eastern corner of the region.

The longitude effect of seasons 2001 to 2003 only differed significantly from one at the edges of the region whilst in the central region it was relatively constant. The longitude effect of season 2004 was well above one in the central region, though this was most likely an artefact of the small effect values on the eastern edge. Therefore, it was decided to remove the longitude effect and only consider the spatial effect due to latitude. The resulting multiplicative median polish model involving main, month, phase and latitude effects (Figure 5.20) had values for each of the effects as seen previously. This confirmed the exclusion of the longitude effect.

Median polish model effects were also computed for the combined seasons. They displayed similarities across all six combined seasons and their features were similar to those seen for the median polish model effects of the individual seasons.

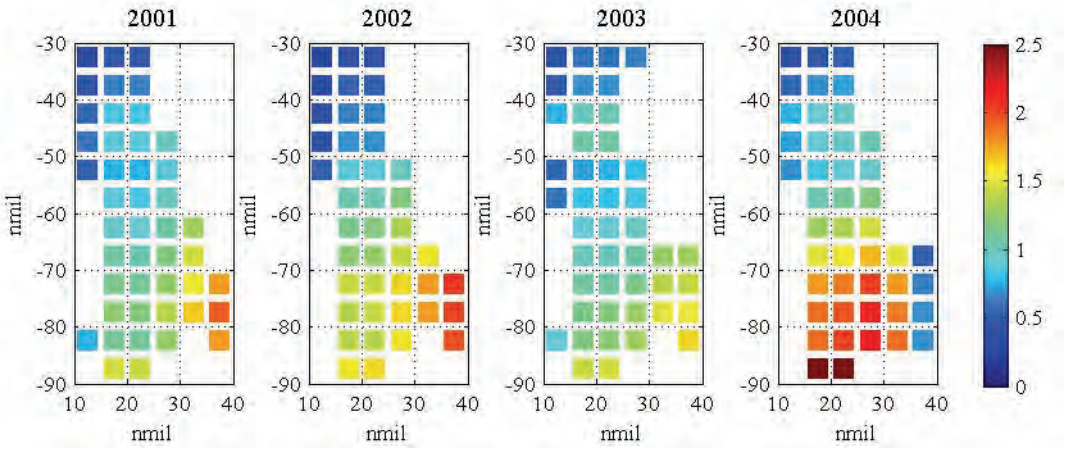


Figure 5.19. Combined spatial effect (latitude and longitude) for individual seasons, 2001-2004.

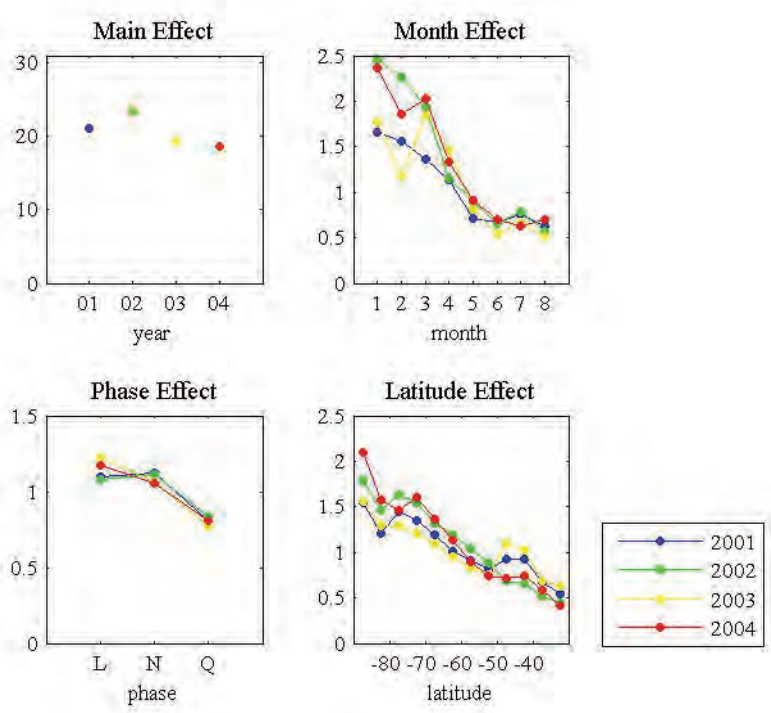


Figure 5.20. Multiplicative median polish spatiotemporal trend effects for individual seasons, 2001-2004.

5.4 COMPARISON OF TREND MODELS

For all individual and combined seasons, the temporal effects (lunar month and phase) of the median polish model are very similar to the temporal trend seen in the classical decomposition model. The latitude effects of the median polish model are quite similar across the seasons suggesting an increase in value for decreasing latitude. To compare the models, the standardised catch rate variable resulting from the classical decomposition temporal trend model and median polish spatiotemporal trend model were compared.

5.5 STANDARDISED CATCH RATE VARIABLE

The classical decomposition temporal trend model and median polish spatiotemporal trend model obtained for the individual and combined seasons can be used to standardise (detrend and deseasonalise) the *CPUEvd* and *CPUEgd* variables by dividing the CPUE values by the relevant model. We call this new variable the standardised CPUE and denote it by *CPUEs*.

Summary statistics of the *CPUEsvd* (Table 5.6) and *CPUEsgd* (Table 5.7) variables for both standardising methods (classical decomposition and median polish) show that both *CPUEs* variables have a mean of approximately one for all individual seasons. A unity mean was expected from the fitting of multiplicative models. The means of the data sets detrended by classical decomposition are less variable across seasons (vd: 0.97-0.99, gd: 1.04-1.05) whilst those detrended by median polish are more variable across seasons (vd: 0.75-1.02, gd: 0.80-1.06). The medians are slightly lower than their associated means, reflecting the positive

skewness of each data set. The variances of the data sets detrended by median polish are lower than the corresponding data set detrended by classical decomposition. The maxima of CPUEs data sets aggregated by vessel (2.97-5.24) are much lower than those aggregated by grid (6.08-11.12). The minima of all data sets are close to zero.

There is no obvious pattern of skewness for individual seasons or detrending methods. However, the skewness of data sets aggregated by vessel (0.91-2.17) is slightly lower than those aggregated by grid with values varying from (1.71-2.96). Kurtosis values, measuring the peakedness of the distribution are also higher for data sets aggregated by grid.

Table 5.6. Summary statistics of daily *CPUEsvd*, CD and MP models.

	2001		2002		2003		2004	
	CD	MP	CD	MP	CD	MP	CD	MP
Mean	0.98	0.93	0.99	0.75	0.99	1.02	0.97	0.78
Median	0.91	0.88	0.89	0.69	0.83	0.91	0.91	0.73
Variance	0.20	0.14	0.30	0.11	0.40	0.28	0.30	0.12
Cv	0.46	0.40	0.55	0.45	0.63	0.52	0.56	0.44
Skewness	1.07	1.05	1.45	2.17	2.04	1.67	1.08	0.91
Kurtosis	2.04	2.53	3.95	15.29	6.16	4.93	1.83	2.75
Min	0.03	0.03	0.03	0.08	0.05	0.08	0.03	0.02
Max	3.70	2.97	4.81	5.41	5.24	4.75	3.85	3.38
Count	3344	3344	3274	3274	3169	3169	2928	2928

Table 5.7. Summary statistics of daily *CPUEsgd*, CD and MP models.

	2001		2002		2003		2004	
	CD	MP	CD	MP	CD	MP	CD	MP
Mean	1.05	0.98	1.05	0.80	1.04	1.06	1.04	0.85
Median	0.94	0.89	0.92	0.73	0.87	0.95	0.93	0.78
Variance	0.36	0.28	0.42	0.20	0.48	0.39	0.41	0.20
Cv	0.57	0.54	0.62	0.55	0.67	0.59	0.61	0.53
Skewness	1.82	2.55	2.24	2.96	2.28	2.35	1.98	1.71
Kurtosis	7.69	23.78	13.17	27.96	9.80	15.53	9.70	7.86
Min	0.04	0.03	0.05	0.04	0.05	0.05	0.03	0.02
Max	8.18	11.12	9.42	9.13	8.14	10.68	8.68	6.08
Count	7981	7981	7205	7205	6422	6422	6116	6116

The aim of standardisation is to produce a data set that enables the comparison of catch rates across the season to be used in a space-time geostatistical model. Standardisation removes the large-scale variability which may involve a mean and variance that varies spatially, temporally or both. It appears that the classical decomposition detrending process has resulted in a standardised variable that is more consistent across seasons. That is, variables from one season can be compared with another season. It is also necessary to consider how the standardised catch rate variable varies across space and time within each season.

5.5.1 VARIATION WITHIN SEASON

The CPUE values displayed a temporal trend across each season, involving a peak at the beginning of the season and a levelling off at a much lower level by the end of the season (Figure 5.21, top row). The variability of the catch rate values are higher in the first few months and significantly lower on the second half of the season. Standardising by classical decomposition results in a variable with a mean that varies about one across each season (centre row). The variability of the catch rate standardised by classical decomposition is more even across the weeks of the season. Spatiotemporal median polish produces a standardised catch rate with mean below one for most weeks of the four individual seasons, with seasons 2002 and 2004 showing slightly lower means than the other seasons. The variability of the median polish standardised variable is lower across the season than the variable standardised by classical decomposition.

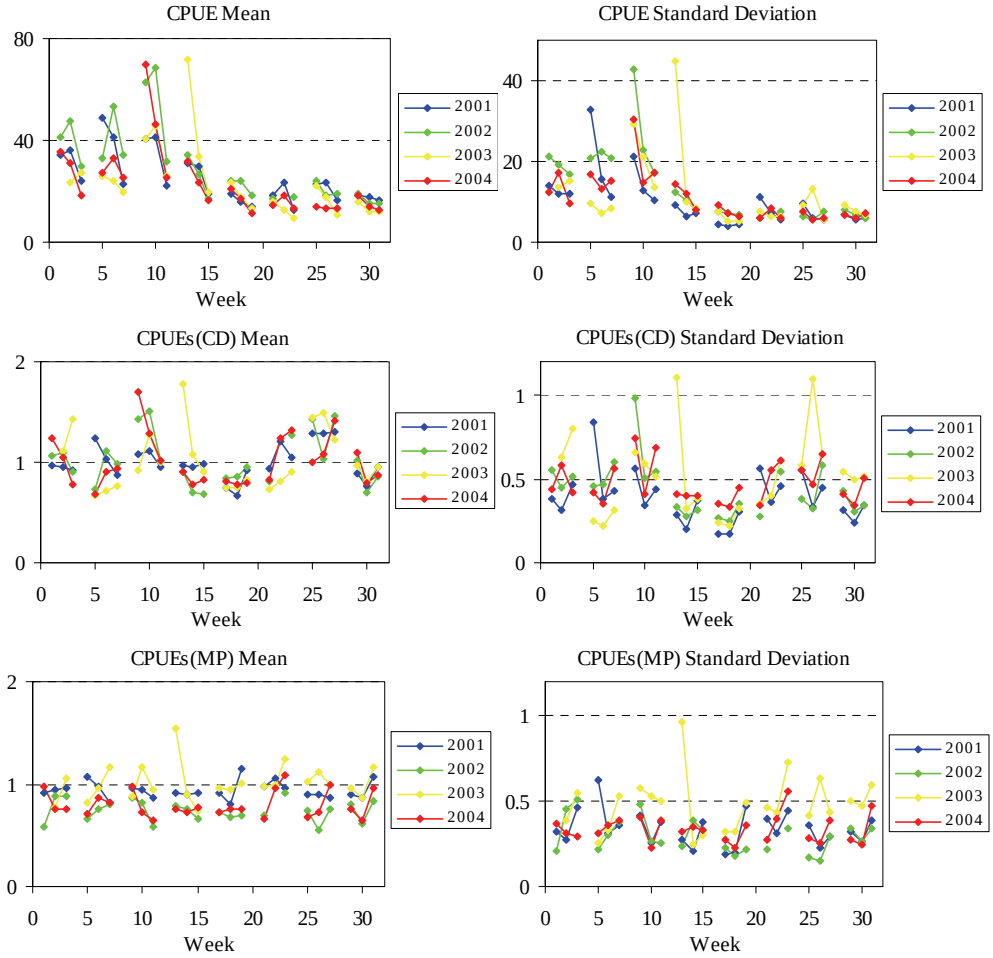


Figure 5.21. Mean and standard deviation of *CPUE* (top row), *CPUEs*(CD) (centre row) and *CPUEs*(MP) (bottom row) values, 2001-2004.

5.5.2 VARIATION ACROSS STUDY REGION

The CPUE values showed a decrease in average value and variability for increasing latitude and increasing longitude (Figure 5.22, top row) for each individual season. Standardising by

classical decomposition does not change these relationships, except that the mean and variability for each season now varies around one and there is reduced variability between the four individual seasons. Standardising by median polish results in a variable whose mean is equal to or less than one, and almost constant for varying latitude within each season. However each season is constant at a different level. For varying longitudes, the catch rate standardised by median polish show greatest variability for the eastern longitude.

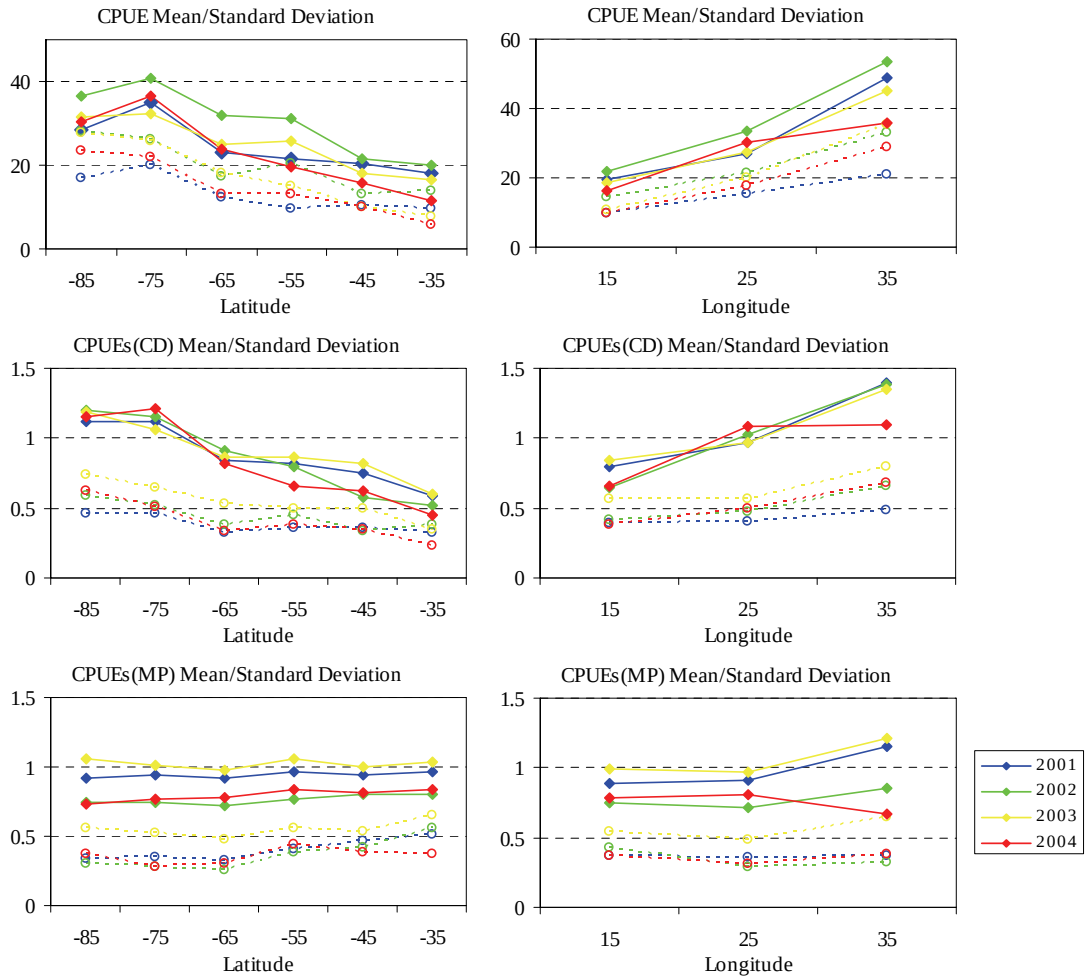


Figure 5.22. Mean and standard deviation of *CPUEs* by latitude and longitude, 2001-2004.

5.6 CHOICE OF STANDARDISATION METHOD

The classical decomposition detrending method produced a standardised catch rate varying about a mean level of one with variability across the season and spatial region varying between zero and one. Means and variances for the individual seasons were quite similar. In contrast, the standardised variable obtained via spatiotemporal median polish detrending showed less variability across the season and region with a mean below one. Whilst the individual seasons showed similarities in mean values and variability across each season, they also displayed

significant differences in their mean values at varying latitudes. Therefore, it was decided to utilise the standardised catch rate obtained via classical decomposition.

The standardisation of the catch rates for a combined season was also investigated by alternative methods which produced almost identical results. The catch rates of an individual season could be standardised by the trend model for the individual season and then combined with the standardised catch rates of other individually standardised seasons to produce a single combined season data set. Alternatively, the parameters of these individual season trend models could be averaged to produce an 'average' model for the combined season, which was then used to detrend all data within the combined season data set. Finally the combined data for all relevant seasons could be used to determine the trend model, which was then used to detrend all records within that data set. In this study, the combined season was detrended by this final method, although the other ways are just as valid.

The aim in standardising catch rates was to produce a model of spatiotemporal continuity of the catch rates which could then be utilised, if appropriate, to forecast the catch rate of a subsequent season. This would produce estimates of the standardised catch rate for the subsequent season, which would then be transformed to catch rate estimates via the global (classical decomposition) trend. As the trend model for a future season is unknown, it was decided to use the trend model obtained for a relevant combined season. The global trend model for the combined season 2001-2003 was chosen as a (estimated) measure of the trend model of the 2004 season based on its similarities with all three individual seasons involved. Moreover these three seasons occur directly before 2004 and with no obvious changes in management for the 2004 season there is no reason to expect a significant change in the trend of the catch rates. The trend model would be used to multiply estimates or simulations of standardised catch rates obtained via geostatistical methods for the 2004 season to produce catch rate values. Furthermore, as the geostatistical estimation and simulation processes to be utilised required conditioning or initialising data from the 2004 season, these conditioning data could be transformed to standardised catch rates using the global trend model of the 2001-2003 combined season as an estimate of the unknown 2004 trend.

5.7 NORMAL SCORES TRANSFORMATION OF *ADJ* VARIABLE

As the geostatistical simulation procedure SGS-ST to be used requires data to be (approximately) normally distributed, the normal scores transformation method was employed

to yield *CPUEsNvd* and *CPUEsNgd* values from the *CPUEsvd* and *CPUEsgd* values that follow a standard normal distribution.

5.8 CHAPTER SUMMARY

Temporal trend models obtained by classical decomposition of the weekly *CPUE* data show great similarity among the individual and combined (consecutive) seasons of 2001 to 2004. A median polish spatiotemporal trend model confirmed the temporal trend of the classical decomposition model. The classical decomposition temporal trend model was used to detrend and deseasonalise the catch rate variable to obtain a standardised catch rate variable with unity mean. This method of standardisation has not previously been applied to the prawn catch rates and makes use of the temporal trend evident in the catch rates over the season. The modelling framework of this study to predict catch rates for the 2004 season will use the temporal trend model of the 2001-2003 seasons as an estimate of the temporal trend for the 2004 season. This is justified by the persistence of temporal trends across the 2001-2003 seasons and the expectation of a similar trend in the 2004 season with no major changes in management.

CHAPTER 6

SPATIAL CORRELATION OVER TIME

6.1 INTRODUCTION

The global trend of the catch rate data has been modelled and used to produce a standardised catch rate to investigate the spatiotemporal continuity of the catch rates across the season. Removal of the global trend allows the spatiotemporal continuity of the catch rate to be modelled without the dominance of the global trend.

Since the goal of spatiotemporal geostatistical analysis is to predict values where no data have been collected, the tools and models will only work on spatiotemporally dependent data. If data are spatiotemporally independent, there is no possibility to predict values at unsampled locations or times. Therefore, before considering the (standardised) catch rates in a spatiotemporal model, it is useful to consider the spatial correlation of the (standardised) catch rates over time to see if there is temporal persistence of the spatial behaviour of the catch rates.

6.2 SPATIAL CONTINUITY OF CATCH RATES WITHIN SEASON

Spatial variography of (lunar) weekly sets of the daily king prawn catch rates (aggregated by vessel) had previously been undertaken and showed that the catch rates displayed spatial continuity (Mueller et al., 2008). Semivariograms of the king prawn *CPUE* were modelled with spatial variograms involving a nugget effect and one spherical structure. The variograms of individual seasons 2001 to 2004 showed high relative nuggets in the weeks corresponding to the new moon lunar week. These weeks also displayed the shortest range.

Standardised catch rate data aggregated daily by either vessel or 1 by 1 nmil grid were grouped by lunar week in order to conduct spatial variogram analysis of the associated *CPUEs* variables. There were 24 weeks with data over 8 lunar months as there were no data in the full moon weeks. The adjusted variables were positively skewed so variograms of their normal scores were also investigated as some geostatistical methods require normally distributed data. Omnidirectional semivariograms of *CPUEsvd, CPUEsgd, CPUEsNvd* and *CPUEsNgd* were computed using the GSLIB routine *gamv*, with an example of the parameter file shown in APPENDIX G.1. No anisotropy was considered due to the shape of the study region.

The standardised experimental semivariograms (semivariogram divided by the variance) for the *CPUEsvd* and *CPUEsgd* catch rate variables of the 2004 season are shown for

comparison (Figure 6.1). There is clear evidence of spatial dependence of the standardised catch rate from their semivariograms. The first five lunar months (March-July) show the clearest evidence of spatial autocorrelation with most of the weeks in the last three lunar months of the season showing much more variability in their spatial structure. The other seasons (2001 to 2003) also show similar evidence of spatial autocorrelation (APPENDIX H.1).

Experimental semivariograms of the *CPUEsvd* and *CPUEsgd* variables for each lunar week show similarities across most of the lunar weeks. This behaviour is also seen in other seasons (APPENDIX H.1). There is a steady increase in variogram value for increasing lag spacing, with the *CPUEsgd* data showing higher variogram values for almost all lag spacings and most weeks. A levelling off of both *CPUEsvd* and *CPUEsgd* variograms is evident between the 10 nmil and 20 nmil lag spacing. The pattern of a steady increase in variogram value for increasing lag spacing combined with a levelling out of the variogram at large lag spacings is indicative of the spatial continuity of the variable. The shape of the semivariograms of the *CPUEsvd* and *CPUEsgd* variables across the 24 weeks of each season are similar suggesting the spatial continuity of the catch rate is both constant and persistent across a season.

The semivariogram values of the *CPUEsgd* variable are larger at lower lag spacings than those for the corresponding *CPUEsvd* variable for all lunar weeks of the individual seasons 2001-2004, . The *CPUEsgd* semivariograms are also much smoother than the corresponding *CPUEsvd* semivariograms. These characteristics can be attributed to the aggregation by grid for the *CPUEsgd* data where the averaging process has inflated the small-scale variability of the data. Overall the semivariograms for the gridded data are less variable than the data aggregated by vessel. The spatial autocorrelation of the data aggregated by vessel is more apparent with more variability evident for increasing spatial separation.

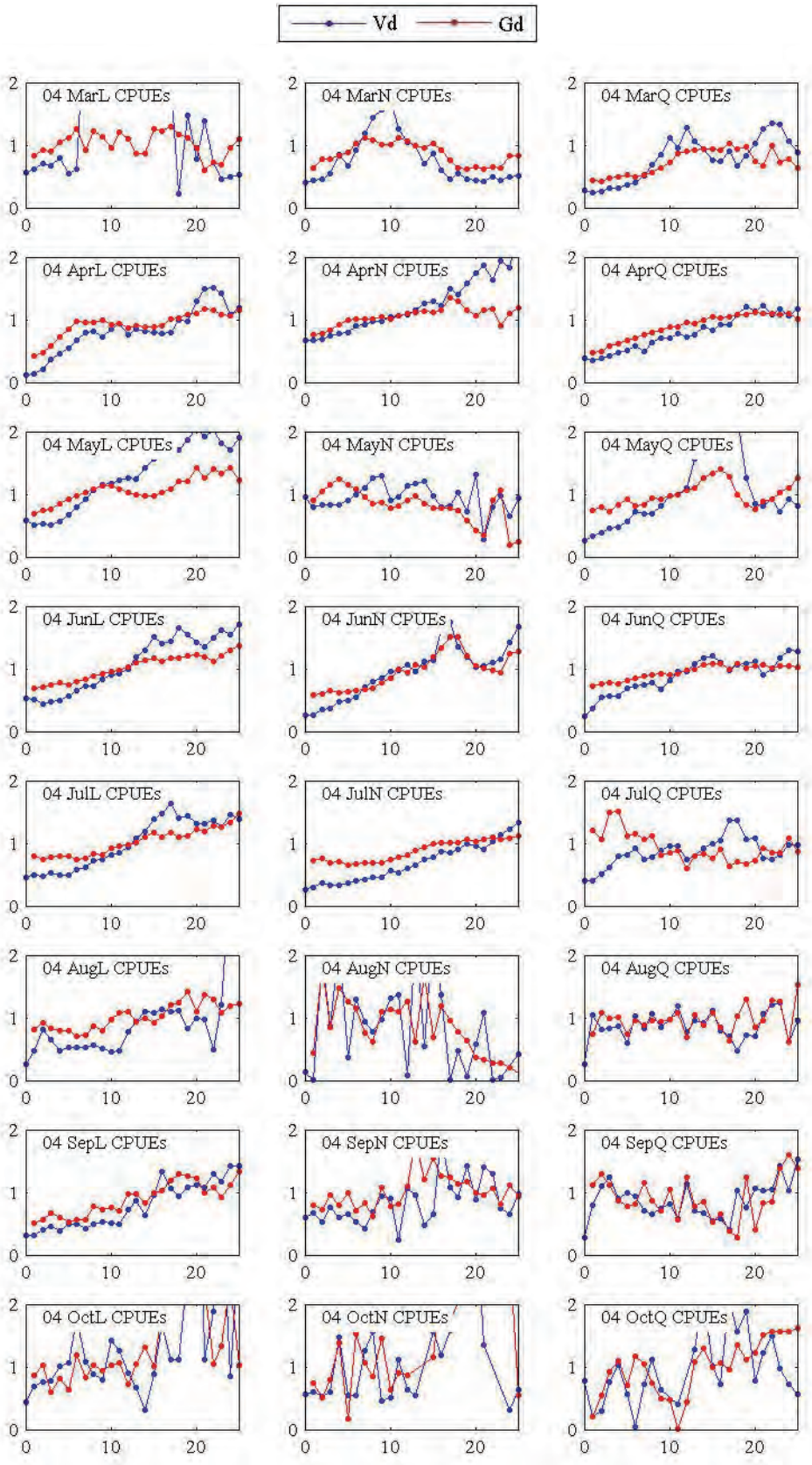


Figure 6.1. Standardised spatial experimental semivariograms of *CPUEsvd* (blue) and *CPUEsgd* (red) for the lunar weeks of 2004.

Consecutive weeks of the standardised catch rates were investigated to examine the persistence of spatial continuity over time. In order to compute spatial correlations of catch rates at various times, the original catch rate values were averaged over a 5 by 5 nmil grid for each (lunar) week, resulting in 48 spatial grid locations. This grid size was used to enable the comparison for pairs of consecutive weeks in a given season as well as for comparative weeks of the four individual seasons.

Tjøstheim's index is a numerical measure to compare if the location of the standardised catch rate ranked *i* in one week is the same as, or close to, the location ranked *i* for the subsequent week (Tjøstheim, 1978). Positive spatial association of the standardised catch rates is seen for many pairs of consecutive weeks in the first half of the four individual seasons (Figure 6.2, top row). However, not all of these quotients are statistically significant at the 5% confidence level. Season 2002 shows the most number of consecutive weeks with significant spatial association. There are few pairs of significant positive correlation for the second half of all seasons.

Spearman's rank correlation was also considered for pairs of consecutive weeks in each season as a measure of the correspondence between the rankings of the spatial location for each week. The rank correlation does not require the dependence between variables to be represented by a linear relationship. The linear correlation coefficient, which is sensitive only to a linear relationship between two variables, is also computed to assess the linear dependence of the catch rates of consecutive weeks. Most of the pairs of consecutive weeks in seasons 2002, 2003 and 2004 showed positive rank and linear correlation (Figure 6.2, middle and bottom rows). Season 2001 showed positive correlations in the first half of the season but mainly negative correlations (rank and linear) in the second half of the season.

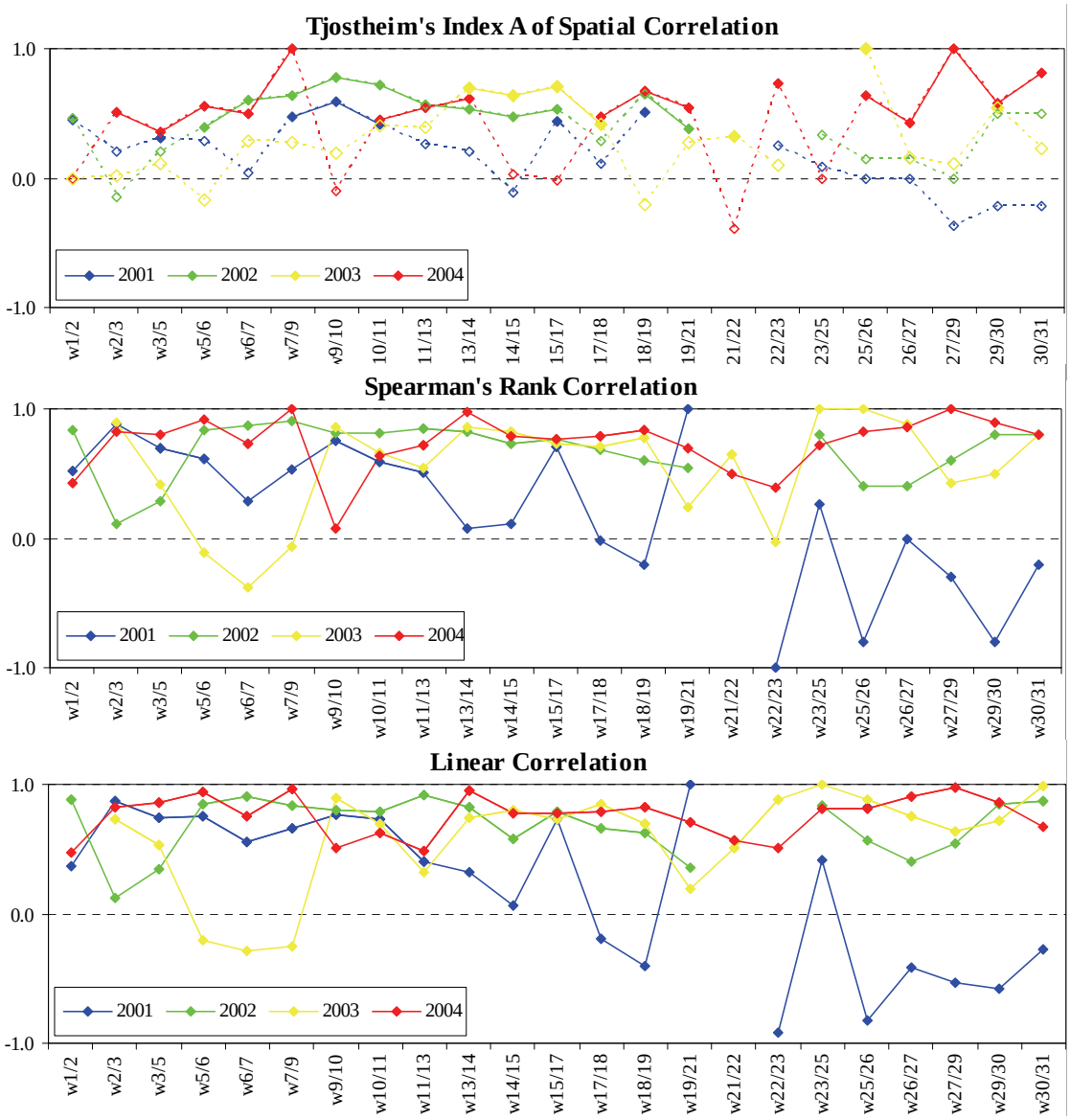


Figure 6.2. Tjostheim's Index of Spatial Correlation (top row), Spearman's Rank Correlation (middle row) and Linear Correlation Coefficient Coefficient (bottom row) of standardised catch rates of adjacent weeks, 2001-2004. Significant Tjostheim's indices shown in bold.

6.3 SPATIAL CONTINUITY OF CATCH RATES ACROSS SEASONS

The spatial structure evident in the semivariograms of the catch rates standardised by vessel (*CPUEsvd*) and grid (*CPUEsgd*) was compared across the individual seasons 2001 to 2004 for each of the lunar weeks. For *CPUEsvd* across the four individual seasons (Figure 6.4) the weeks of the New Moon and First Quarter show great similarity for the four seasons. The Last Quarter weeks (left column) of the first four months shows the greatest variation between seasons as one or more seasons in each of these weeks displays significantly higher variogram values for all lag spacings. This is attributed to the number of high catch rate values evident in these weeks as the Carnarvon-Peron Line and Extended Nursery Area are opened to fishing. As noted previously, the variograms of the last three lunar months show little structure for all seasons. This is most

likely due to the small number of data available for each of these weeks. In a similar manner, the spatial structure of the lunar weekly semivariograms of *CPUEsgd* from corresponding weeks of each season showed consistent behaviour across the four individual seasons (Figure 6.5).

An analysis of Pearson's, rank and spatial correlation was undertaken to assess the relationship of catch rates of comparative weeks between seasons (Figure 6.3). As this required similar spatial locations over the various weeks, catch rate data aggregated on a 5 nmil by 5 nmil grid were used. Many positive correlations were seen in the weeks of the first half of each season. Correlations for the second half of each pair of seasons were more variable, with both positive and negative correlations observed. This was partially due to the lack of data in many of the spatial grid locations considered.

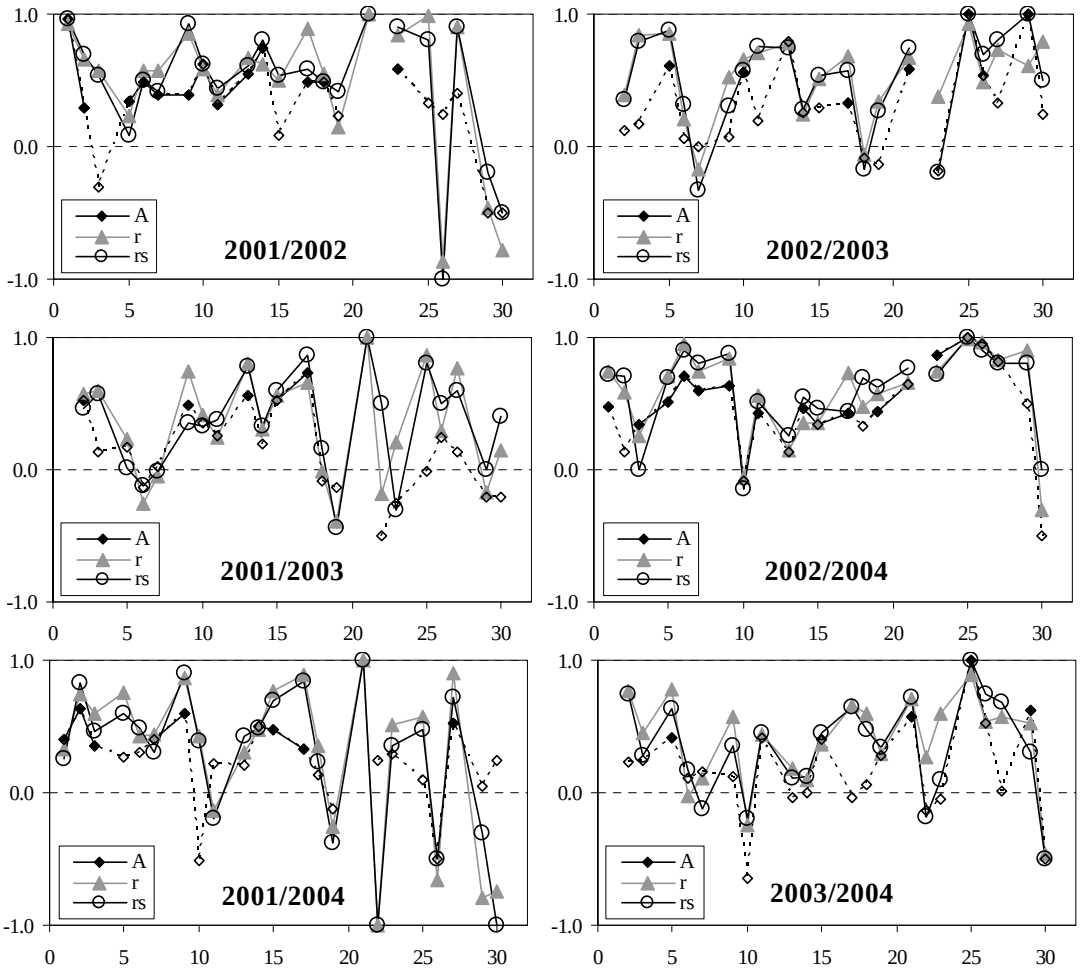


Figure 6.3. Tjostheim's Index of Spatial Correlation (A), Spearman's Rank Correlation (r_S)and Pearson's Correlation Coefficient Coefficient (r) of standardised catch rates aggregated by 5 nmil block for comparative weeks in seasons 2001-2004. Significant Tjostheim's indices shown in bold.

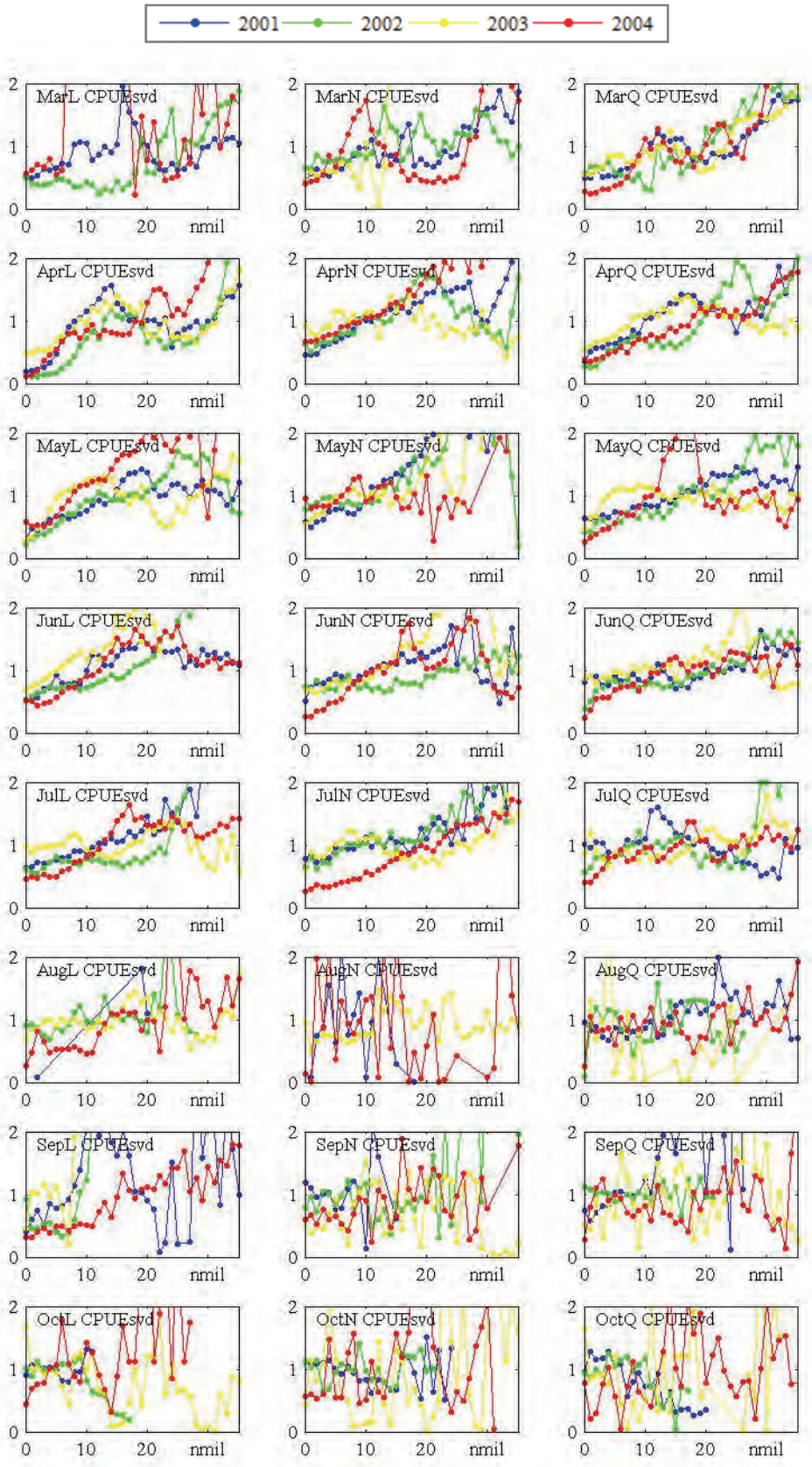


Figure 6.4. Standardised spatial experimental semivariograms of *CPUEsvd* for the lunar weeks of the individual seasons 2001-2004.

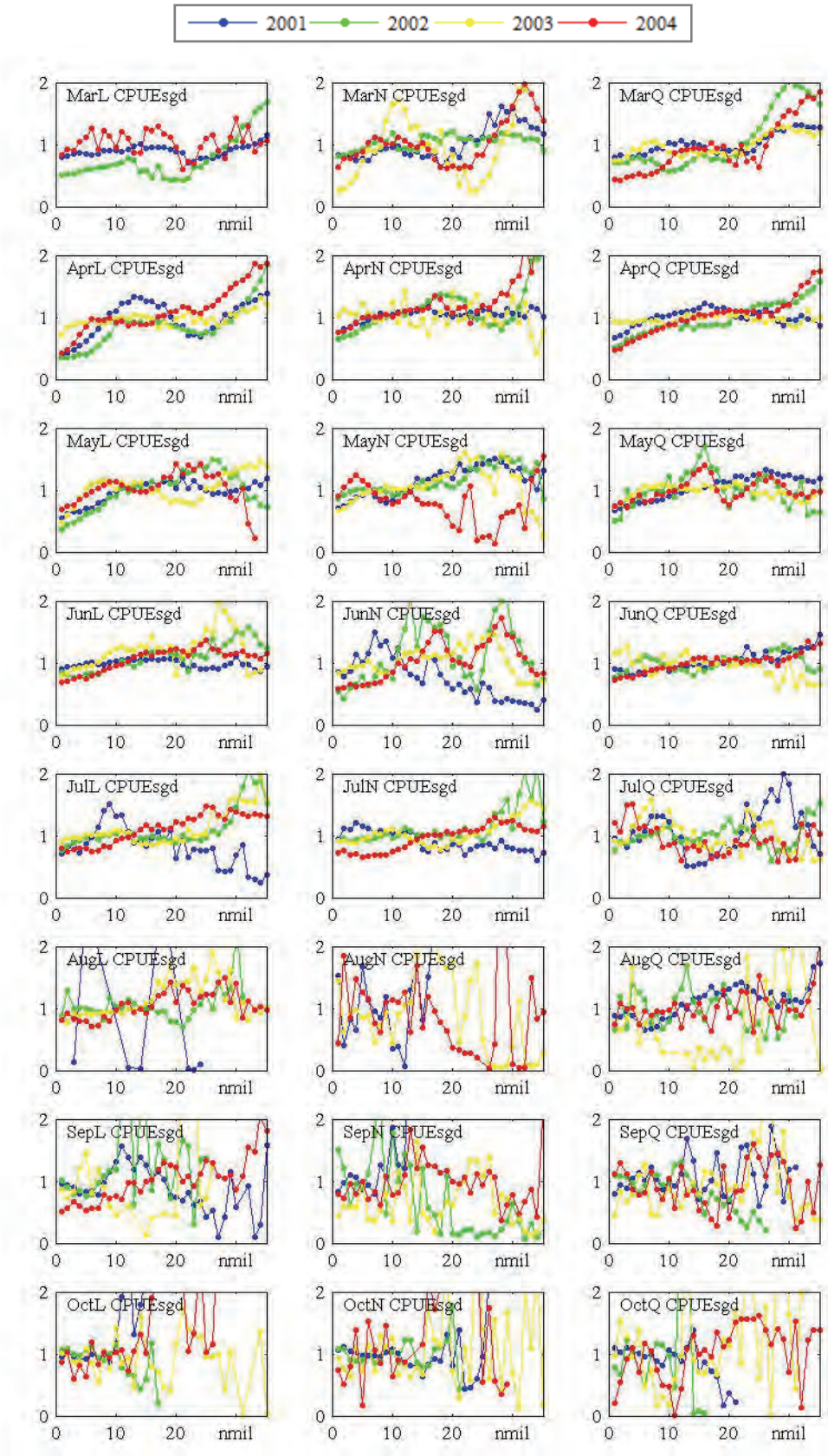


Figure 6.5. Standardised spatial experimental semivariograms of *CPUEsgd* for the lunar weeks of the individual seasons 2001-2004.

6.4 SPATIAL VARIOGRAM MODELLING

For the weeks displaying evidence of spatial autocorrelation, the experimental semivariogram was modelled by a semivariogram typically involving a nugget and one spherical structure. A few weeks also required a second spherical structure. The semivariogram model parameters fitted to the experimental semivariograms of the *CPUEsvd* and *CPUEsgd* variables are shown in APPENDIX H.2.

The relative nugget of the models of the *CPUEsvd* data sets are smaller in the first few months of the season with many models increasing to almost pure nugget in the final weeks of the season (Figure 6.6, left). The *CPUEsgd* data sets also show an increase in relative nugget across the season (Figure 6.6, right). The relative nuggets of the *CPUEsvd* data sets are generally smaller at the beginning of the season than those of the *CPUEsgd* data sets.

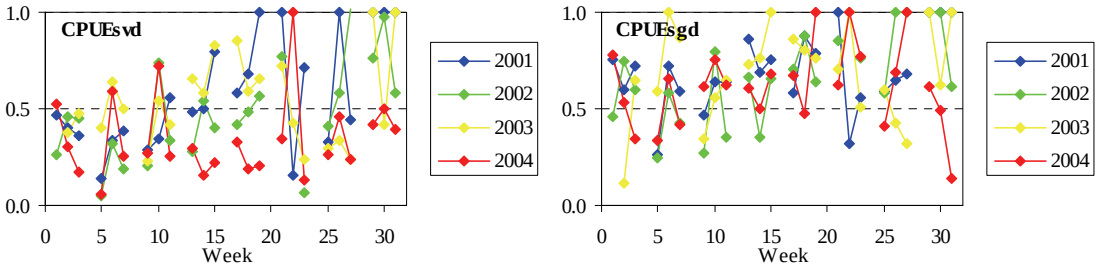


Figure 6.6. Relative nugget of fitted spatial semivariogram models of *CPUEsvd* (left) and *CPUEsgd* (right) for the lunar weeks of the individual seasons 2001-2004.

The range of the semivariogram is the distance in which the difference of the variogram from the sill becomes negligible. The ranges seen in the semivariogram models for both the *CPUEsvd* and *CPUEsgd* variables vary across the season from 2 to 30 nmil (Figure 6.7).

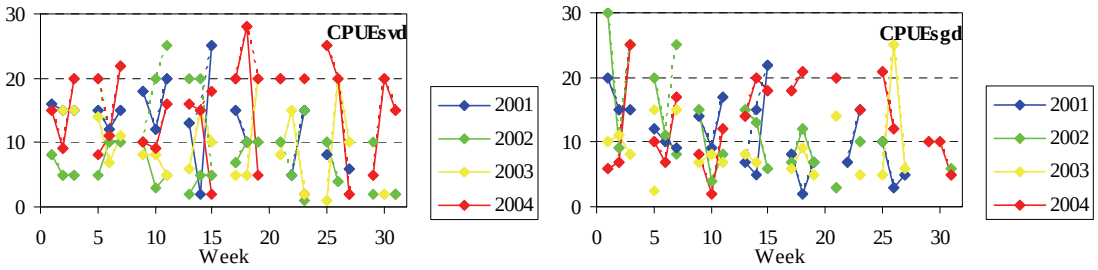


Figure 6.7. Range of fitted spatial semivariogram models of *CPUEsvd* (left) and *CPUEsgd* (right) for the lunar weeks of the individual seasons 2001-2004.

The relative nuggets of the *CPUEsvd* and *CPUEsgd* variables for a given season vary across the season (Figure 6.8). The catch rates aggregated by vessel consistently show smaller relative nuggets than those aggregated by grid for the same week. This was observed in the experimental semivariograms (Figure 6.1) and can be attributed to the support effect due to

aggregation over a 1 nmil by 1 nmil grid, which removes the small-scale spatial variability of the data. Comparing ranges for the two methods of catch rate aggregation (Figure 6.9) shows similarities between the two with no method demonstrating an obvious pattern of smaller or greater ranges.

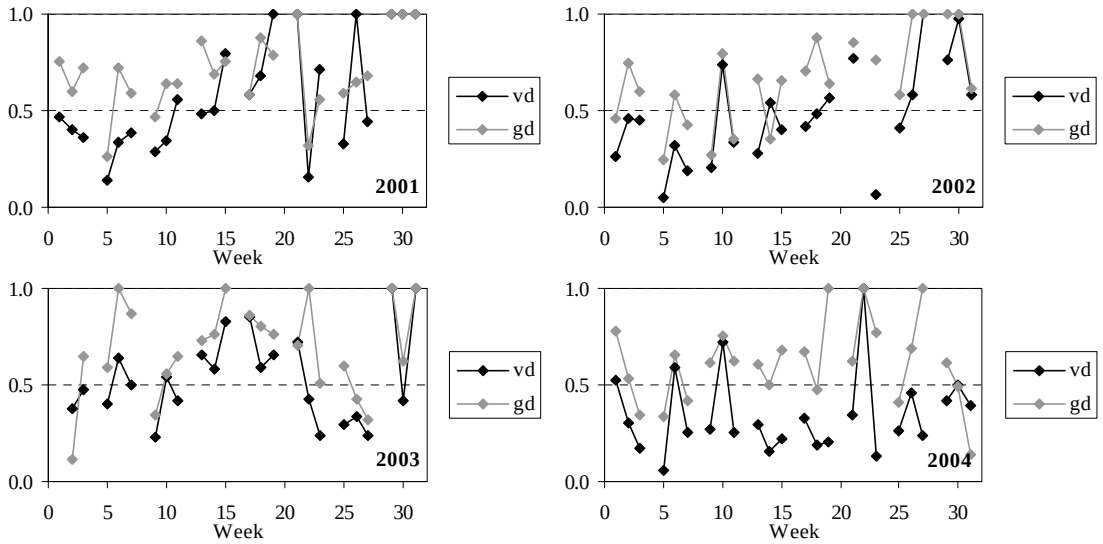


Figure 6.8. Relative nugget of fitted semivariogram models of *CPUEsvd* (black) and *CPUEsgd* (grey) for the lunar weeks of the individual seasons 2001-2004.

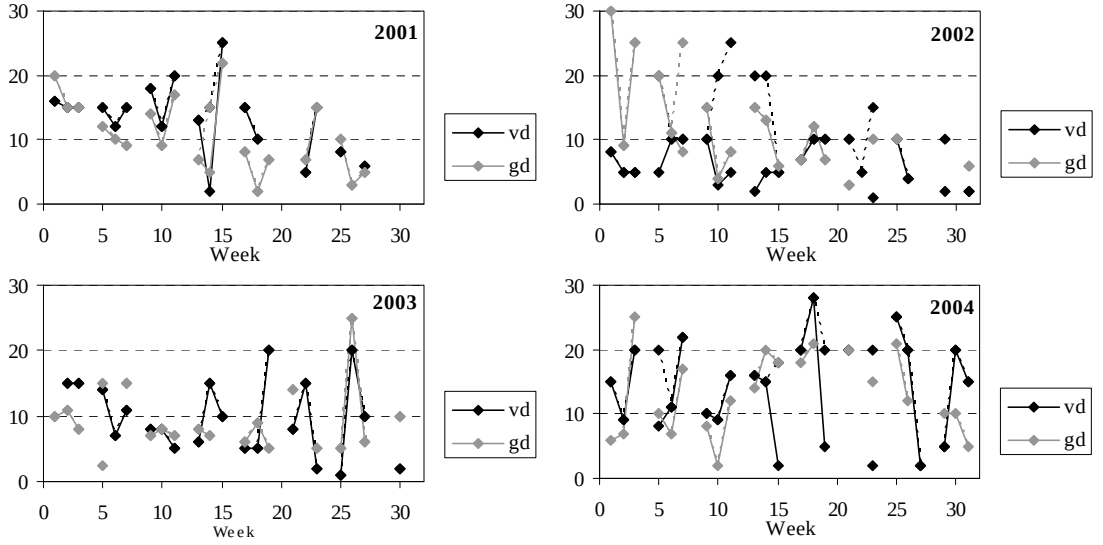


Figure 6.9. Range of fitted semivariogram models of *CPUEsvd* (black) and *CPUEsgd* (grey) for the lunar weeks of the individual seasons 2001-2004.

6.5 SPATIAL CONTINUITY OF NORMAL SCORES

The evidence of structure seen in the adjusted data is also present in the spatial semivariograms of the associated normal score variables (Figure 6.10). The spatial structure of the normal scores of each lunar week is very similar. The variogram values of the normal scores of the data aggregated by grid location show slightly higher values for small lag spacings than the data

aggregated by vessel. The spatial structure of *CPUEsvdN* and *CPUEsgdN* across the four individual seasons is similar to that seen for the standardised catch rates (APPENDIX H.3).

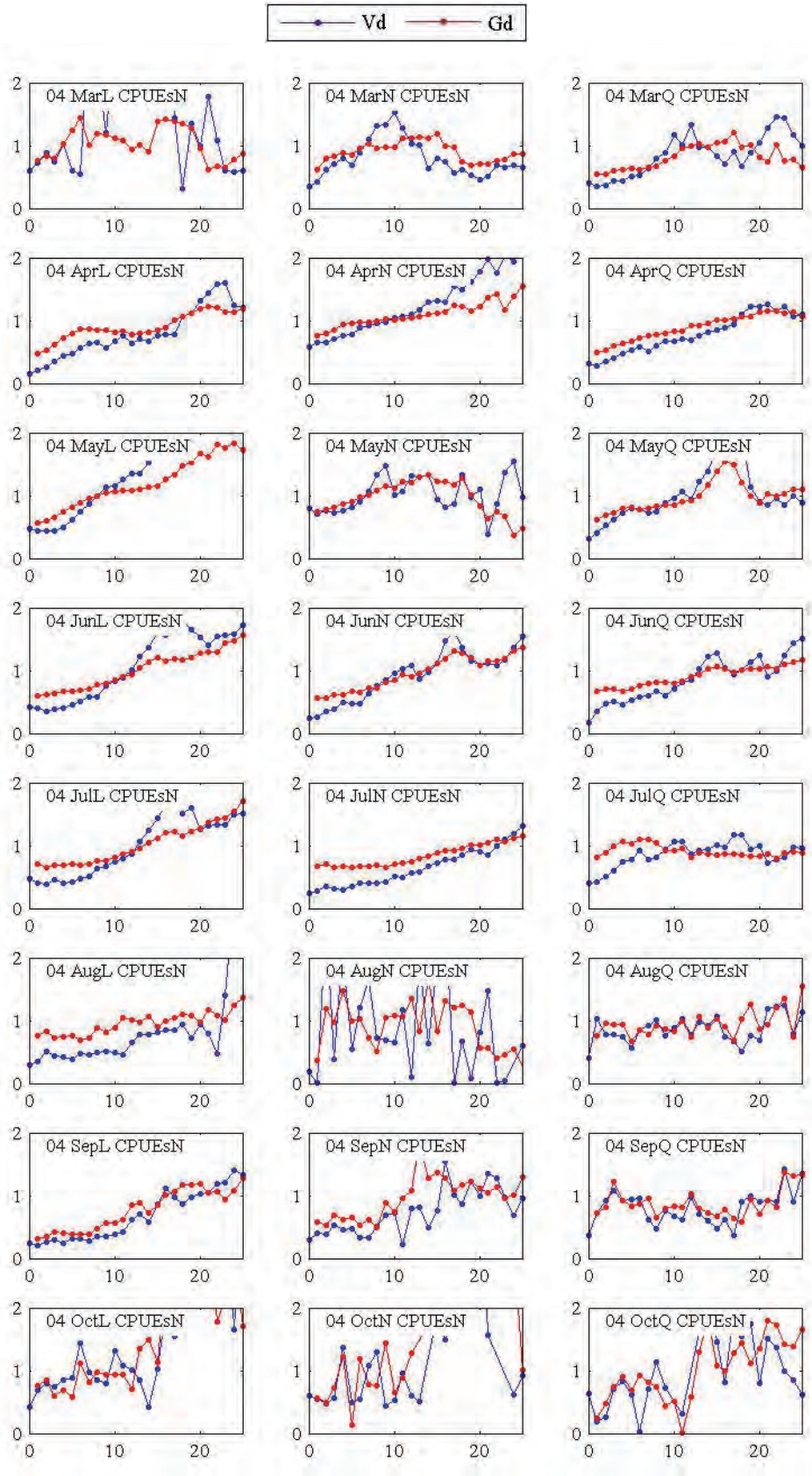


Figure 6.10. Standardised spatial semivariograms of *CPUEsvdN* (blue) and *CPUEsgdN* (red), lunar weeks of 2004.

6.6 CHAPTER SUMMARY

Moderate positive spatial correlation is evident between the standardised catch rates (aggregated by vessel or grid) in adjacent weeks of all four individual seasons 2001-2004. Weak to moderate positive spatial correlation is also seen between comparable weeks in different seasons. Spatial continuity of standardised catch rates is evident in most lunar weeks of each season. Moreover, the pattern of spatial continuity, as measured by semivariogram, is similar throughout each season and across the various seasons. Semivariograms of standardised catch rates aggregated by vessel show a smaller discontinuity due to randomness at the origin than the catch rates aggregated by grid. For the data aggregated by grid the larger semivariogram values for separation distances of up to 10 nmil are attributed to the support effect resulting from the aggregation process. The temporal persistence of spatial continuity within each season and across the seasons is indicative of spatiotemporal autocorrelation of the standardised catch rates and supports the analysis via spatiotemporal geostatistical methods.

CHAPTER 7

SPATIOTEMPORAL VARIOGRAPHY

7.1 INTRODUCTION

The spatial continuity of weekly data sets and the temporal persistence of their behaviour, seen in the previous chapter, supports the use of spatiotemporal geostatistical methods to analyse the spatiotemporal continuity of the king prawn catch rate data. Modelling of the spatiotemporal continuity of the standardised catch rate is presented in this chapter by computing and modelling spatiotemporal semivariograms. As outlined previously, the modelling framework of this study requires a model of spatiotemporal continuity for the 2004 season based on the 2001-2003 seasons. This model of spatiotemporal continuity can then be used to facilitate geostatistical spatiotemporal estimation and simulation of the king prawn catch rate data. The spatiotemporal variography of the 2004 season is shown in this chapter for comparison and completeness but will not be used in the subsequent estimation and simulation processes.

7.2 SPATIOTEMPORAL VARIOGRAPHY OF INDIVIDUAL SEASONS

Spatiotemporal semivariograms were computed to characterise the spatiotemporal continuity of the standardised catch rate variable as a measure of the dissimilarity of data pairs as the (spatial and/or temporal) separation between them increases. Experimental semivariograms were calculated for each individual season (2001-2004) using a temporal lag spacing of 1 day and a spatial lag spacing of 0.5 nmil after investigating various temporal and spatial lags. The number of lags was chosen to allow a maximum temporal lag of up to 30 days and a maximum spatial lag up to 30 nmil. Directional and omnidirectional semivariograms were considered for the spatial plane. The standardised catch rate values, and their normal scores, of both daily aggregation methods were considered for spatiotemporal variography. The experimental semivariograms were computed using the GSLIB *gamvmod.exe* program, with parameter file as shown in APPENDIX G.2, and post-processing in MATLAB.

7.2.1 EXPERIMENTAL SPATIOTEMPORAL VARIOGRAMS

Experimental spatiotemporal semivariograms for the *CPUEsvd* data of the individual seasons (Figure 7.1) demonstrate the presence of spatiotemporal continuity. In the absence of spatiotemporal autocorrelation, the mathematical expectation of the semivariogram is the

sample variance and the variogram would appear flat. Omnidirectional semivariograms were first considered for the spatial direction. The individual seasons show great similarity, with less evidence of spatial structure in season 2003. There is a relatively stable increase in value with increasing temporal lag and increasing spatial lag suggesting spatiotemporal autocorrelation that is higher for short spatial distances and small temporal separations, and decreases as the distance and time between data points increases. The maximum value of the colour scale is set to the variance of each data set to enable comparison of spatiotemporal continuity between seasons. It must be remembered that their variances of the *CPUEsvd* data vary between 0.200 and 0.395 (Table 7.1).

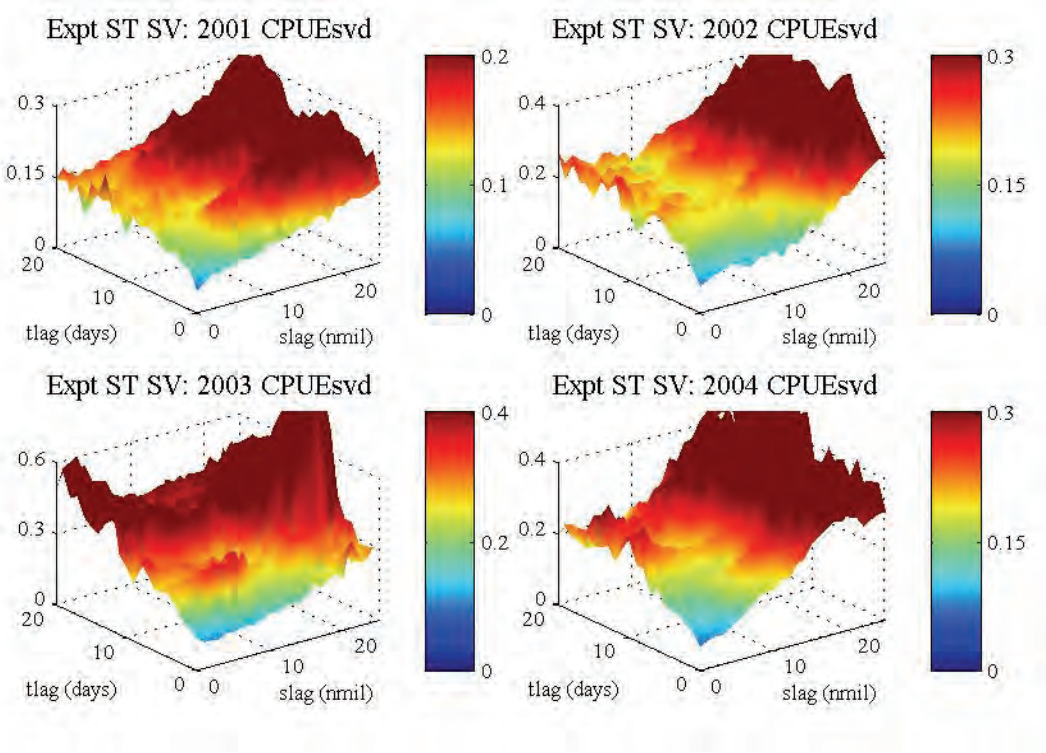


Figure 7.1. Experimental spatiotemporal semivariograms of *CPUEsvd* for individual seasons 2001-2004.

Table 7.1. Standardised catch rate variance for individual seasons, 2001-2004

Season	2001	2002	2003	2004
CPUEsvd	0.200	0.297	0.395	0.296
CPUEsgd	0.423	0.453	0.587	0.414

The range of the semivariogram is interpreted as the distance beyond which pairs are no longer spatially or temporally correlated. At this distance, the variogram plateaus at a value known as the sill. The temporal range of the four individual seasons is between 10 to 15 days, and the spatial range is between 10 and 20 nmil. Investigation of the spatiotemporal

semivariograms in all seasons beyond these ranges show a levelling off in the temporal direction. However, in the spatial direction, the semivariogram values increase and fluctuate quite dramatically. The semivariogram eventually returns to the sill level around 40 nmil separation (not shown), which is almost the maximum spatial separation distance for the data. A possible explanation for this peak in variability, between 20 and 40 nmil separation, before returning to the sill is the use of data pairs from two distinct fishing areas in the north and south of the region. Whilst these regions may exhibit similar spatiotemporal continuity within their own region, the use of data pairs from the two ends of the fishery may not be a reasonable assumption of stationarity. This amounts to a decision of quasi-stationarity for spatial separation less than 20 nmil (Armstrong, 1998).

The experimental semivariograms of the four seasons show a discontinuity at the origin, referred to as the nugget component of the variance, which represents the measurement error or microscale variability of the variable at distances shorter than the smallest distance among samples. All nugget effects are less than 0.1. The nugget accounts for approximately 25-30% of the total variability given by the overall sill. This variability at small spatial and temporal separations indicates that the fine-scale variability of the catch rates is not fully explained by the semivariogram models. This variability can be partially attributed to the aggregation of catch rates.

Semivariograms of the *CPUEsgd* data (Figure 7.2) also show increasing value with increasing temporal and spatial lags, with more erratic structure in season 2003. The variances of the data aggregated by grid are higher than those aggregated by vessel (Table 7.1) reflecting the higher values evident in their experimental spatiotemporal semivariograms. The *CPUEsgd* semivariograms show much higher relative nugget effects and are smoother than those of the *CPUEsvd* data. This is attributed to the smoothing process of aggregating catch and effort data over a 1 by 1 nmil grid. When data are averaged over space and/or time there is a smoothing of the regionalised variable as the space or time window becomes larger (Chiles & Delfiner, 1999).

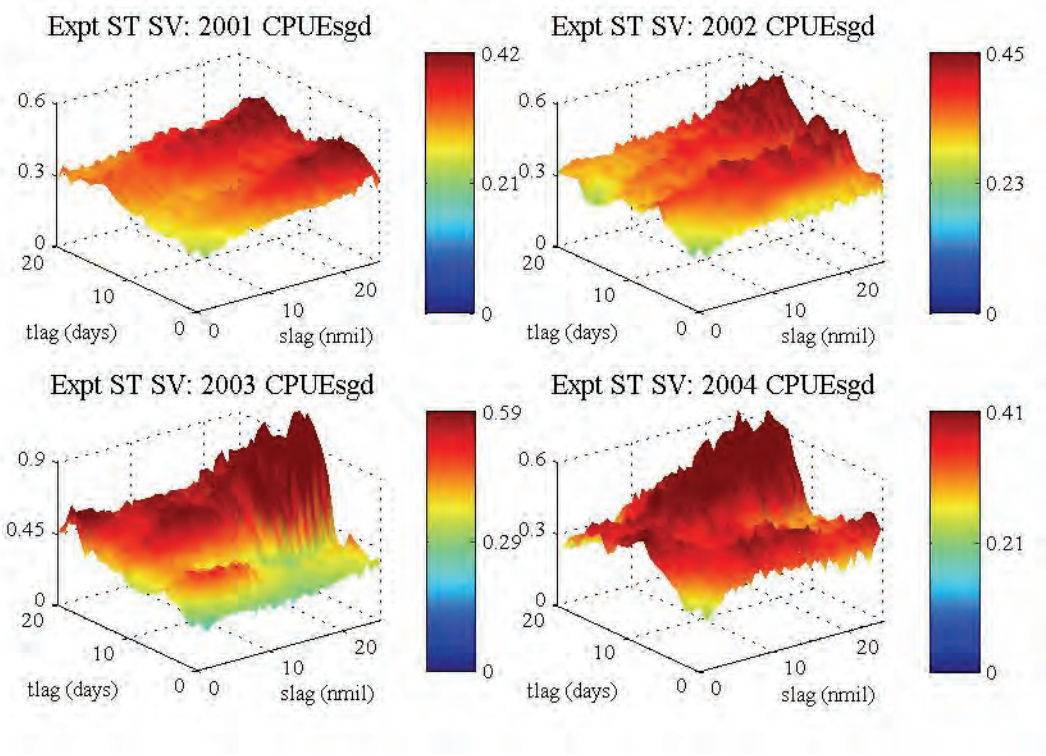


Figure 7.2. Experimental spatiotemporal semivariograms of *CPUEsgd* for individual seasons 2001-2004.

7.2.1.1 DIRECTIONAL SEMIVARIOGRAMS

Directional spatial semivariograms were considered to investigate the possibility of anisotropy. A variable is said to be spatially anisotropic when its pattern of spatial variability changes with direction (Goovaerts, 1997). Two directions were considered, the first was the North-South (NS) direction and the second was the East-West (EW) direction, including a 45° angular tolerance on each side of the specified direction. These directions were chosen due to the span of the fishing region covered by the catch rate data. There exists a relatively small range of separation distances in the East-West direction with a maximum separation distance of around 20 nmil. In contrast, in the North-South direction, the span of the fishing region resulted in separation distances up to 60 nmil, although the level of spatial correlation had reached its plateau at these larger separation distances between catch rates at the northern and southern extremes of the fishery.

The semivariogram for the NS direction of the *CPUEsvd* data set (Figure 7.3) showed a similar pattern of spatiotemporal behaviour to that exhibited by the region as a whole (Figure 7.1). However, the EW direction revealed a more erratic pattern of variability (Figure 7.4). Comparison of the omnidirectional semivariogram with those of the two directions considered shows that all three semivariograms are similar for spatial separation distances less than 15 nmil and temporal separations less than 15 days. The largest observable differences occur for spatial separations greater than 15 nmil, regardless of temporal separation. This spatial separation

occurs when one of the data pairs is within, and the other is outside, the ENA. Higher catch rates typically occur within the ENA resulting in an increased measure of dissimilarity with data located outside the ENA. This results in higher semivariogram values for these data pairs, and explains the semivariogram seen for the EW direction. These directional semivariograms support the use of the omnidirectional semivariogram as a measure of spatiotemporal continuity of the catch rates, and the assumption of quasi-stationarity for spatial separation distances less than 15 nmil.

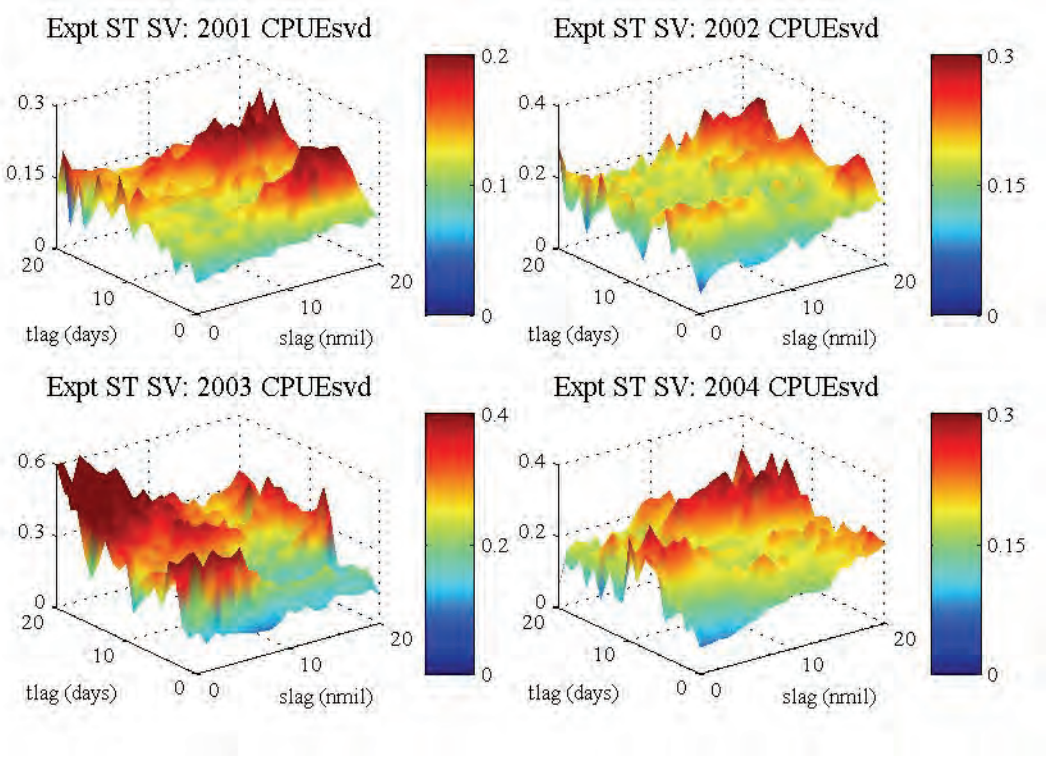


Figure 7.3. NS Directional spatiotemporal semivariograms of *CPUEsvd* for individual seasons 2001-2004.

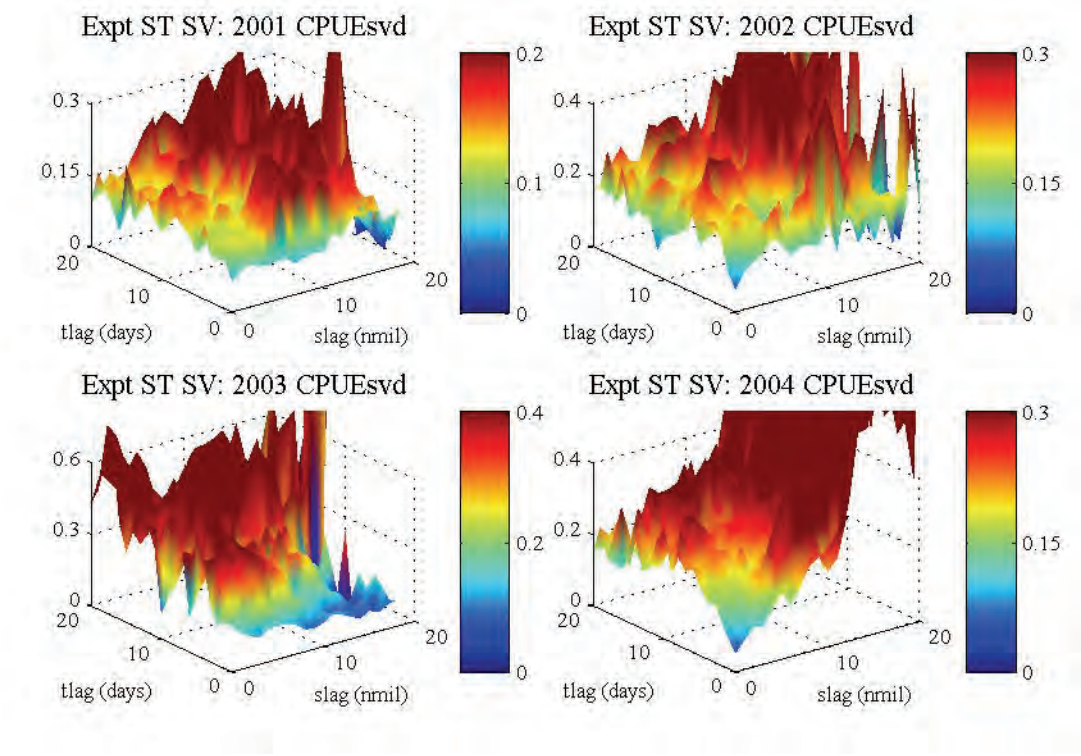


Figure 7.4. EW Directional spatiotemporal semivariograms of *CPUEsvd* for individual seasons 2001-2004.

7.2.1.2 RELATIVE SEMIVARIOGRAMS

Heteroscedasticity is a change in the local variability of data over the study area. The proportional effect is a particular form of heteroscedasticity where the local variance of data is related to the local mean. For positively skewed data sets, such as the king prawn catch rates, the local variance increases with the local mean. When combined with preferential sampling of high-valued areas, due to targeted fishing, a proportional effect may make the sample semivariogram erratic as the clustering of high values means that the data pairs that contribute to small lags come mainly from high-valued areas. The corresponding lag mean is large and the lag variance is also large. As the separation distance increases, the data that contribute to the lag become more representative of the entire area. This trend in lag variance results in overestimating the semivariogram value at short lags and hence also its relative nugget effect.

The general relative semivariogram scales each semivariogram value by a function of the lag mean, typically the squared lag mean is used for positively skewed distributions. The pairwise relative semivariogram scales each lag increment by the squared average of the head and tail values, which reduces the influence of each large value. Relative semivariograms are not substitutes for the traditional semivariogram but they can provide a clearer picture of the spatiotemporal continuity, such as the relative nugget effect and the range. In a spatial-only setting they have shown to reveal spatial structure (correlation) better when data are sparse,

skewed or clustered (Deutsch & Journel, 1998). Due to the denominators applied in each case, relative semivariograms are limited to use with strictly positive variables.

The general relative semivariogram of the standardised catch rate aggregated by vessel (Figure 7.5) is almost identical to its traditional semivariogram (Figure 7.1) whilst its pairwise relative semivariogram reveals a much smoother pattern of spatiotemporal continuity (Figure 7.6). Similarly, the relative semivariogram of the *gd* data (Figure 7.7) reveals a much clearer picture of the spatiotemporal continuity than its traditional semivariogram (Figure 7.2). The pairwise relative semivariogram of the *gd* data has removed much of the variability evident in the traditional semivariogram (Figure 7.8) indicating little evidence of spatiotemporal continuity. The general relative or pairwise relative semivariograms of the *CPUEsgd* and *CPUEsvd* data sets can be used to help infer the semivariogram model parameters including nugget effect, range and type of structure (eg, spherical or exponential).

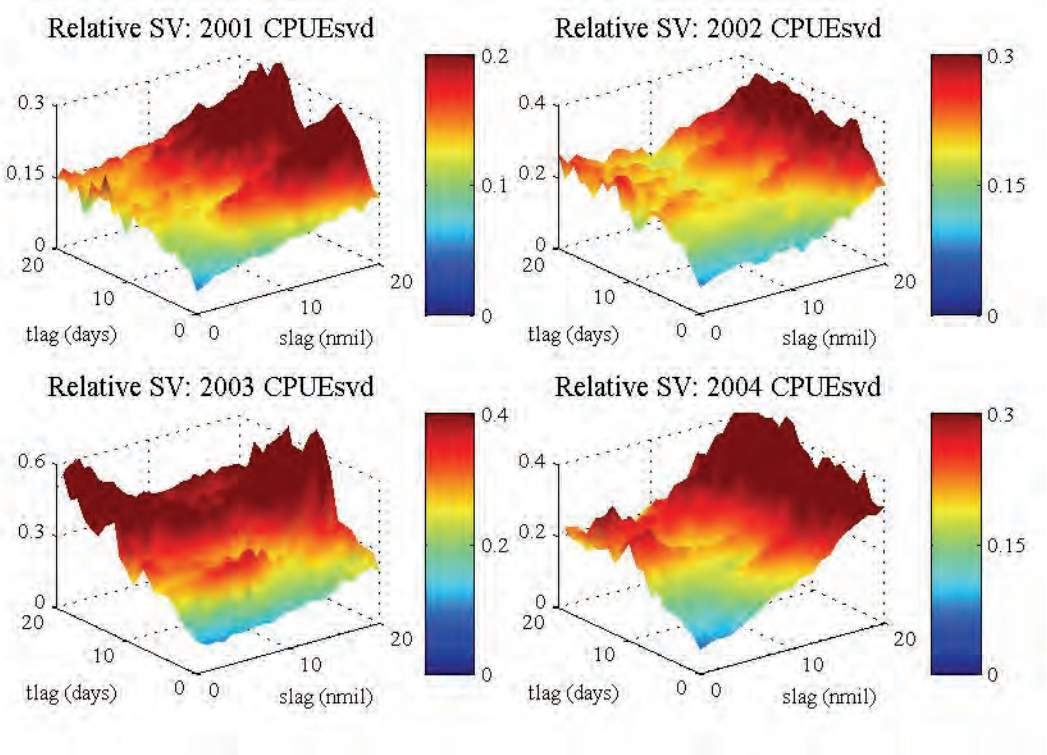


Figure 7.5. General relative semivariograms of *CPUEsvd* for individual seasons 2001-2004.

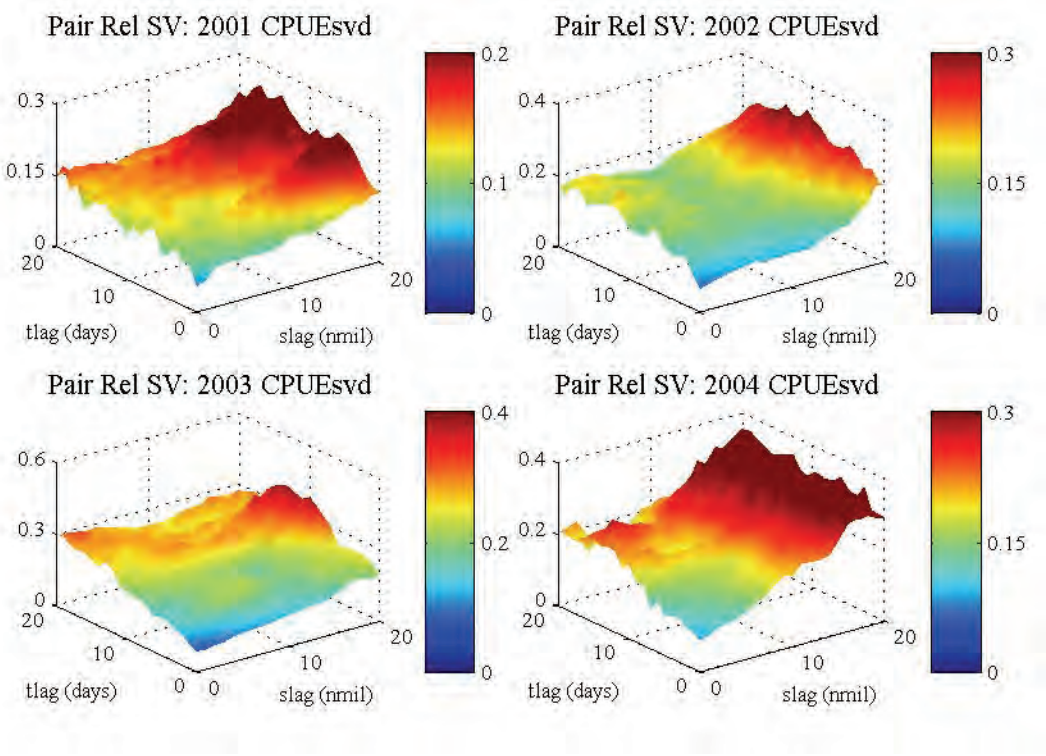


Figure 7.6. Pairwise relative semivariograms of *CPUEsvd* for individual seasons 2001-2004.

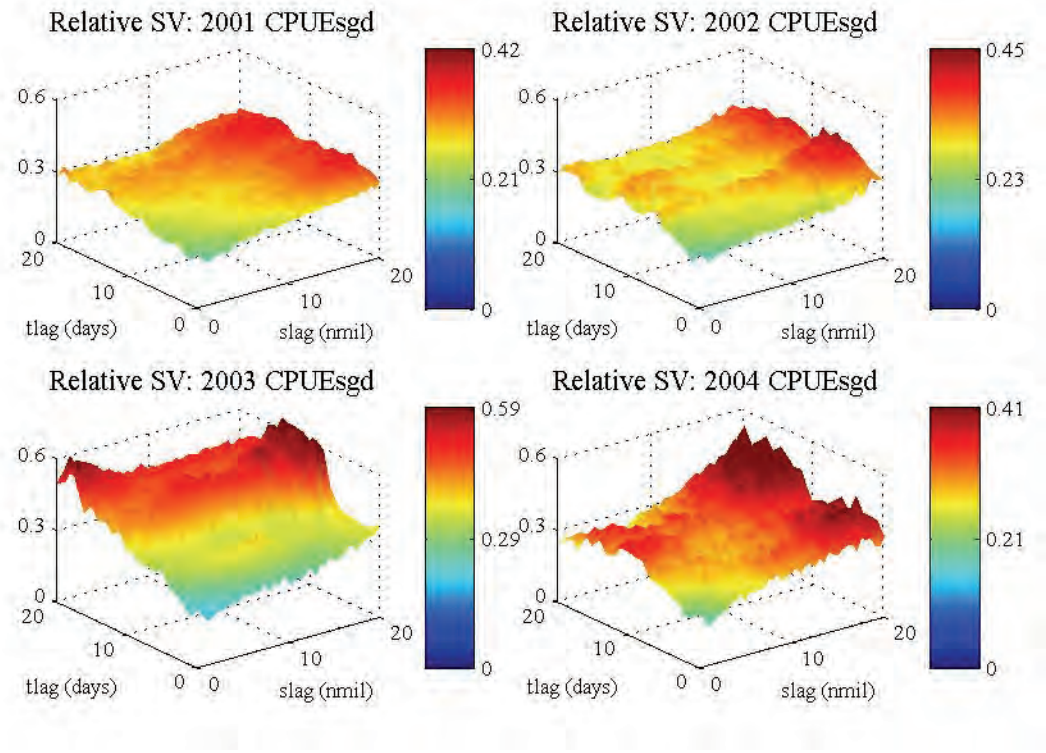


Figure 7.7. General relative semivariograms of *CPUEsgd* for individual seasons 2001-2004.

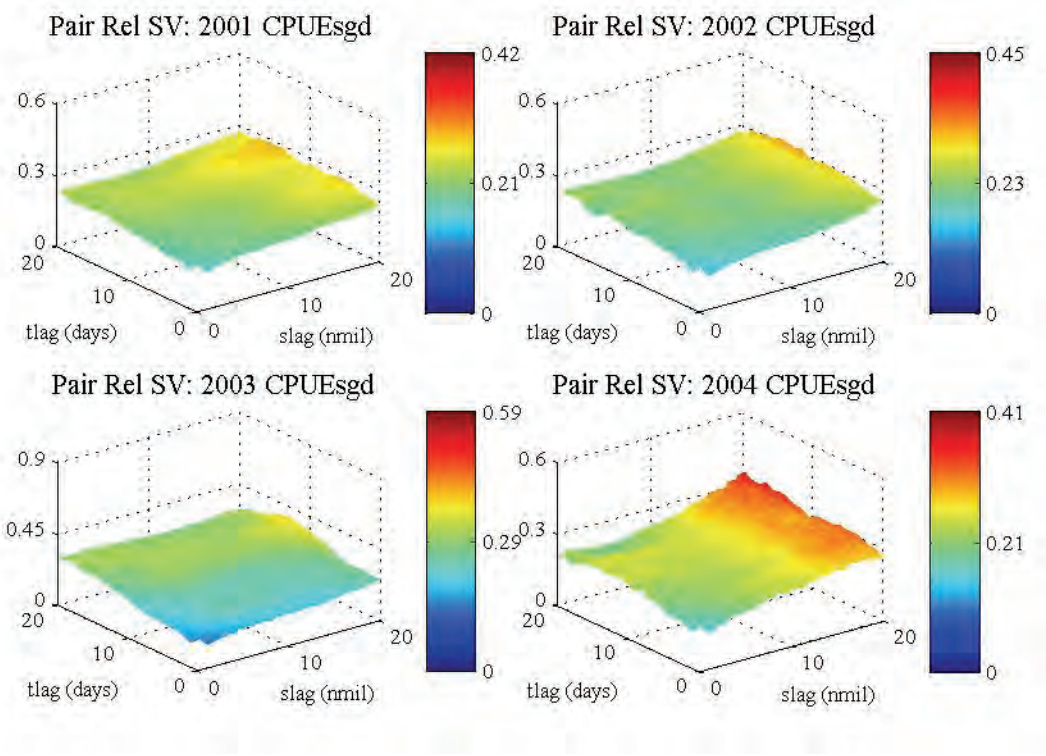


Figure 7.8. Pairwise relative semivariograms of *CPUEsgd* for individual seasons 2001-2004.

7.2.1.3 SEMIVARIOGRAM OF THE LOGARITHMS

The semivariogram of logarithms is the traditional semivariogram computed on the natural logarithms of the original variables. It requires that the original values are strictly positive. Logarithmic semivariograms are considered for use in a lognormal kriging process. This is a nonlinear kriging algorithm which amounts to applying a linear kriging process (SK or OK) to the nonlinear log transform of the data.

The logarithmic semivariogram of the *vd* data is similar to its traditional variogram, with the most noticeable difference in season 2003. The appearance of the logarithmic semivariogram did not show any significant improvement over the traditional semivariogram and therefore was not considered further for modelling of the spatiotemporal continuity of the catch rates. Although it will not be considered further in this study, it is recognised that data are often log-transformed in fisheries modelling, and therefore it is recommended that the use of log-transformed catch rate data is considered in future research of this method.

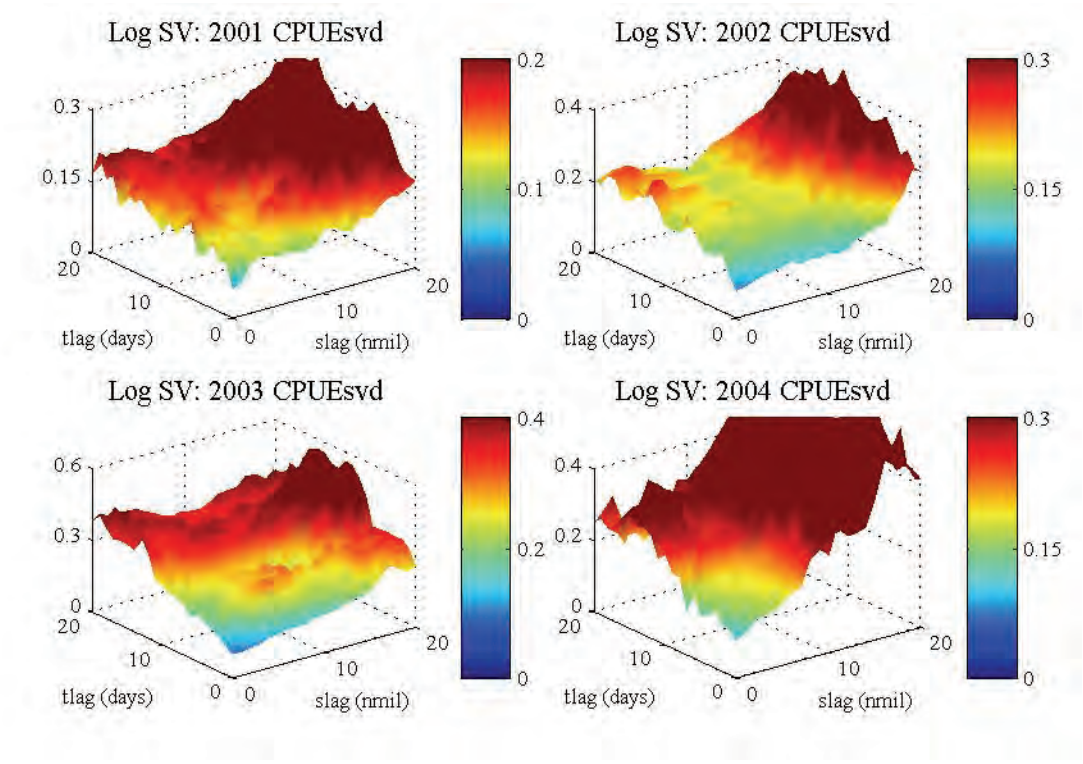


Figure 7.9. Experimental semivariograms of logarithms of *CPUEsvd* for individual seasons 2001-2004.

7.2.2 MODELLING OF SPACE-TIME CONTINUITY

Having established the pattern of spatiotemporal continuity by calculation of the experimental semivariogram, modelling of the semivariogram (in continuous terms) is necessary for geostatistical estimation and simulation so that any spatiotemporal lag can be considered. A model also allows inference to be drawn on various properties of the modelled relationship, such as its (effective) range and asymptotic variance (sill). The semivariogram can be modelled with any conditionally negative definite function. Whilst it is possible to propose a model and test its permissibility, in practice it is customary to use one of a set of basic models that are known to be permissible. This is a relatively straightforward practice in modelling spatial variograms as there are many such models in common use and these models may be combined linearly to form complex models. However, the use of spatiotemporal semivariogram models is less common, although it has become an active area or research in the past decade (Cressie & Huang, 1999; De Iaco et al., 2001; Gneiting, 2002; De Iaco, 2010). Tools for fitting the space-time variogram to experimental variograms remain poorly developed (Heuvelink & Griffith, 2010).

Whilst automatic fitting procedures can be used to achieve an optimal set of parameters, the choice of model to be fitted and the constraints of their parameters make this procedure difficult. Visual inspection of the experimental spatiotemporal semivariogram and its associated

marginal spatial and temporal variograms can reveal the most suitable model to be used and the basic structures to be included. For the catch rate data, the level of noise evident in the experimental semivariograms made automatic fitting problematic. Therefore, the choice of model and parameters was undertaken though a manual fitting process to ensure the salient features of the experimental semivariograms were captured by the semivariogram models. The relative semivariograms were used to help infer the model parameters.

The metric, linear, product and product-sum models were considered as semivariogram models. The metric and product model required a global sill equal to the maximum of the spatial and temporal sills. In addition the metric model assumes the same structure in all directions. The linear and product-sum models allowed different forms and sill in the temporal and spatial directions. The global sill of the linear model is determined by the spatial and temporal sills whilst the product-sum model has more flexibility for the overall sill. All four models require investigation of the marginal spatial and temporal semivariograms.

7.2.3 MARGINAL SEMIVARIOGRAMS

Marginal semivariograms can be extracted from the relevant experimental spatiotemporal semivariogram. The marginal spatial semivariogram is the semivariogram obtained for zero temporal lag spacing. Similarly, the marginal temporal semivariogram is the semivariogram obtained for zero spatial lag spacing. In practice an almost zero (<0.5 nmil) spatial lag spacing is used if the data are not gridded, as is the case for the *vd* data.

Marginal semivariograms for the standardised data aggregated by vessel (Figure 7.10, top) show similar structure between the four individual seasons, although season 2003 shows much more variability in its temporal semivariogram. The marginal spatial semivariograms show a nugget effect of around 0.1 and then a gradual increase in values until they level out at a range of about 20 nmil. The nugget effect for the temporal semivariograms is around 0.1 and a range of around 10 days is evident.

For the data aggregated by grid (Figure 7.10, bottom) the spatial variograms show very similar structure across the seasons for each year. The relative nugget of both spatial and temporal variograms appears higher than for the data aggregated by vessel. Whilst the range of the temporal variograms appears similar, the range of the spatial variograms is shorter than for the data aggregated by vessel.

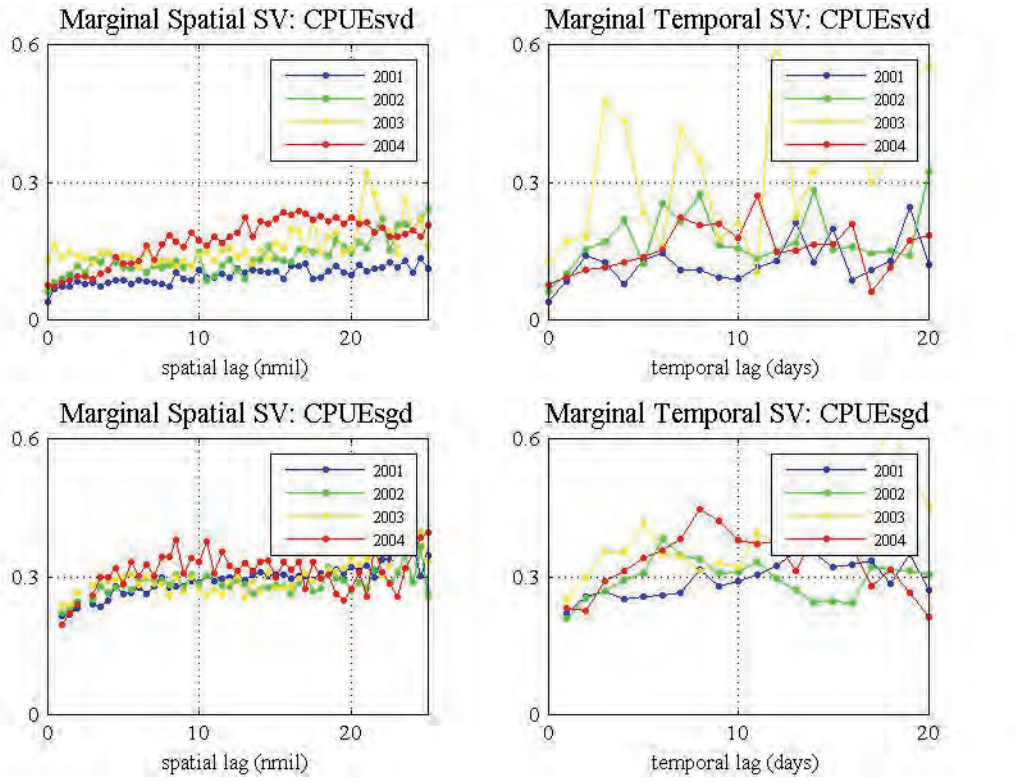


Figure 7.10. Marginal spatial and temporal semivariograms of *CPUEsvd* and *CPUEsgd* for the individual seasons 2001-2004.

The number of pairs for the marginal spatial semivariograms (Figure 7.11) is similar for all seasons of the *vd* data with a peak at the 2 nmil lag spacing. For the *gd* data, the peak occurs at the 4 nmil lag spacing. Similarly the number of pairs for the marginal temporal semivariograms has a peak at the 1 day lag spacing for both methods of aggregation. After these peaks occur there is steady decline in the number of pairs associated with each lag spacing. The peaks of the number of pairs are higher for the *gd* data than the *vd* data.

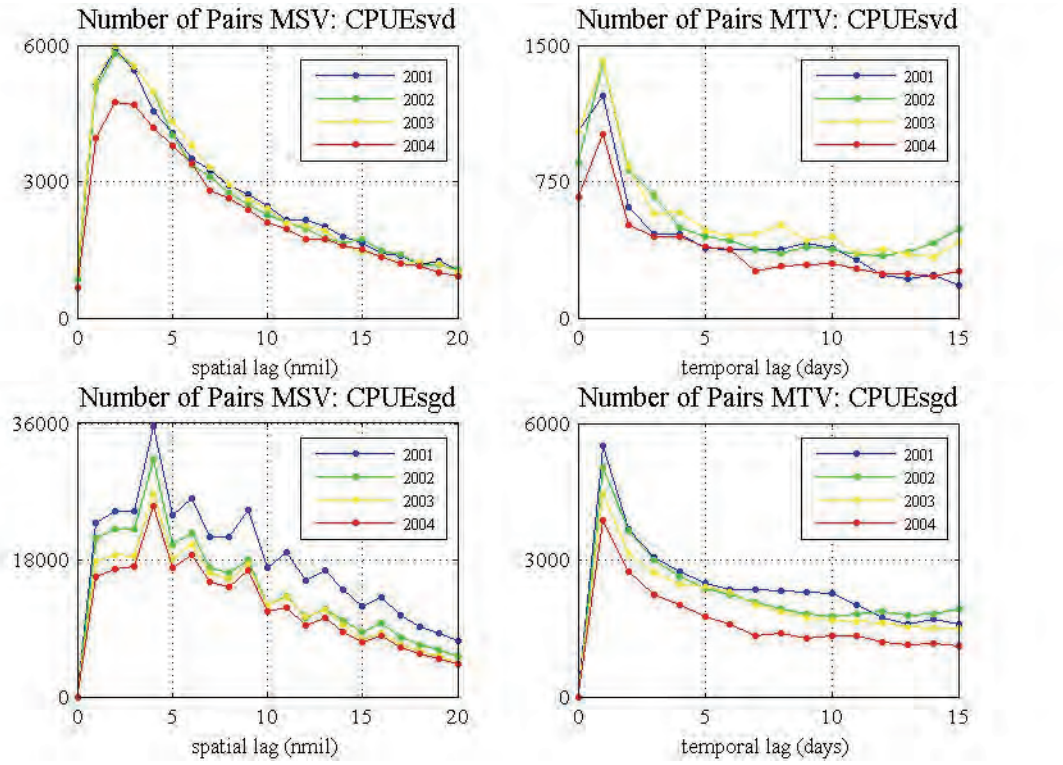


Figure 7.11. Number of pairs for marginal semivariograms of *CPUEsvd* and *CPUEsgd* for individual seasons 2001-2004.

Variogram models were fitted to the marginal spatial and temporal (traditional) variograms of the *vd* and *gd* catch rate data (Table 7.2). The models were fitted manually to reduce the effect the outliers and to take into account the pattern of continuity seen previously in the experimental semivariograms. For instance, in the 2003 marginal spatial variogram of the *vd* data, the nugget effect appeared quite high but at subsequent temporal lags it was lower.

Although the marginal semivariogram model parameters were fitted manually to the experimental semivariograms, it is appropriate to consider statistical criteria to justify the choice of a particular model. Nugget effects were included in all models. Exponential and spherical structures were compared by the weighted sum of squares (Table 7.3). In some cases two spherical structures were required. Temporal lags up to 15 days and spatial lags (0.5 nmil interval) up to 15 nmil were used in the calculations of the weighted sum of squares. Weights used were the number of pairs for the lag standardised by the total number of pairs to enable comparison between the models for the *vd* and *gd* catch rate data. The variogram models for season 2003 showed the poorest fits. The similarities between each season confirm the persistence of spatial and temporal continuity in the catch rates.

Table 7.2. Marginal spatial and temporal semivariogram model parameters of *CPUEsvd* and *CPUEsgd* for individual seasons 2001-2004.

Season	2001 vd	2001 gd	2002 vd	2002 gd	2003 vd	2003 gd	2004 vd	2004 gd
Spatial model								
Nugget	0.04	0.17	0.06	0.18	0.07	0.18	0.06	0.15
1st structure								
Type	sph	exp	sph	exp	sph	exp	sph	exp
Range	1	18	2	15	20	13	10	15
Sill	0.025	0.12	0.03	0.09	0.1	0.09	0.05	0.16
2nd structure								
Type	sph		sph				sph	
Range	15		20				20	
Sill	0.03		0.03				0.1	
Temporal model								
Nugget	0.04	0.17	0.06	0.18	0.12	0.18	0.07	0.15
1st structure								
Type	exp	sph	exp	sph	exp	sph	sph	sph
Range	5	15	8	8	10	6	10	12
Sill	0.09	0.12	0.16	0.15	0.17	0.15	0.1	0.19

Table 7.3. Weighted sum of squares (Equation 2.41) of marginal spatial and temporal semivariogram models of *CPUEsvd* and *CPUEsgd* for individual seasons 2001-2004. Minimum value for fit of each variogram is underlined.

	Marginal spatial model				Marginal temporal model			
	CPUEsvd		CPUEsgd		CPUEsvd		CPUEsgd	
Season	sph	exp	sph	exp	sph	exp	sph	exp
2001	0.009	0.014	0.007	0.002	0.079	0.054	0.012	0.016
2002	0.043	0.044	0.012	0.006	0.147	0.084	0.013	0.031
2003	0.083	0.112	0.002	0.004	0.413	0.356	0.016	0.031
2004	0.007	0.036	0.025	0.017	0.032	0.039	0.020	0.027

After consideration of the weighted sum of squares and visual inspection of the model fits, marginal spatial semivariograms of the *CPUEsvd* data were fitted with models involving a nugget and two spherical structures (Figure 7.12), whilst those of the *CPUEsgd* data only involved a single spherical structure in conjunction with the nugget effect (Figure 7.13). The *CPUEsgd* temporal models involved a nugget and one spherical structure (Figure 7.14) whilst the *CPUEsvd* temporal models involved a nugget and one exponential structure, except for the spherical structure of season 2004 (Figure 7.15). General relative and pairwise relative semivariograms were used to help infer the relative nugget and range of the semivariogram. The models of spatial and temporal continuity are based on a view of the overall continuity evident in each season.

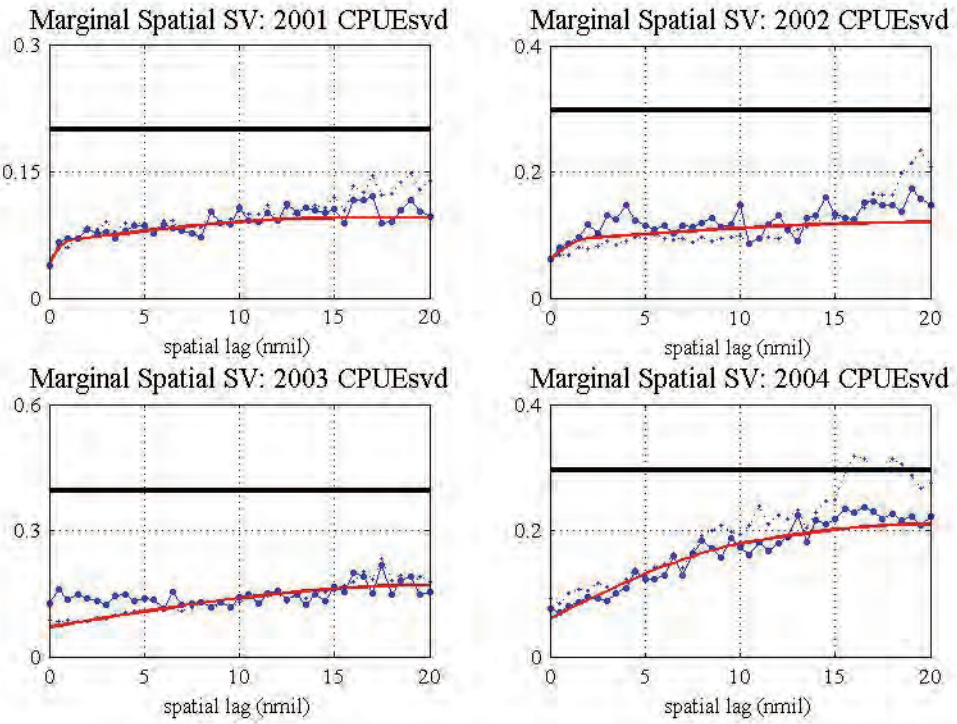


Figure 7.12. Marginal spatial semivariograms (blue) and fitted semivariogram models (red) of *CPUEsvd* for individual seasons 2001-2004. Pairwise relative semivariogram shown by dotted blue line. Variance of data set shown by solid black line.

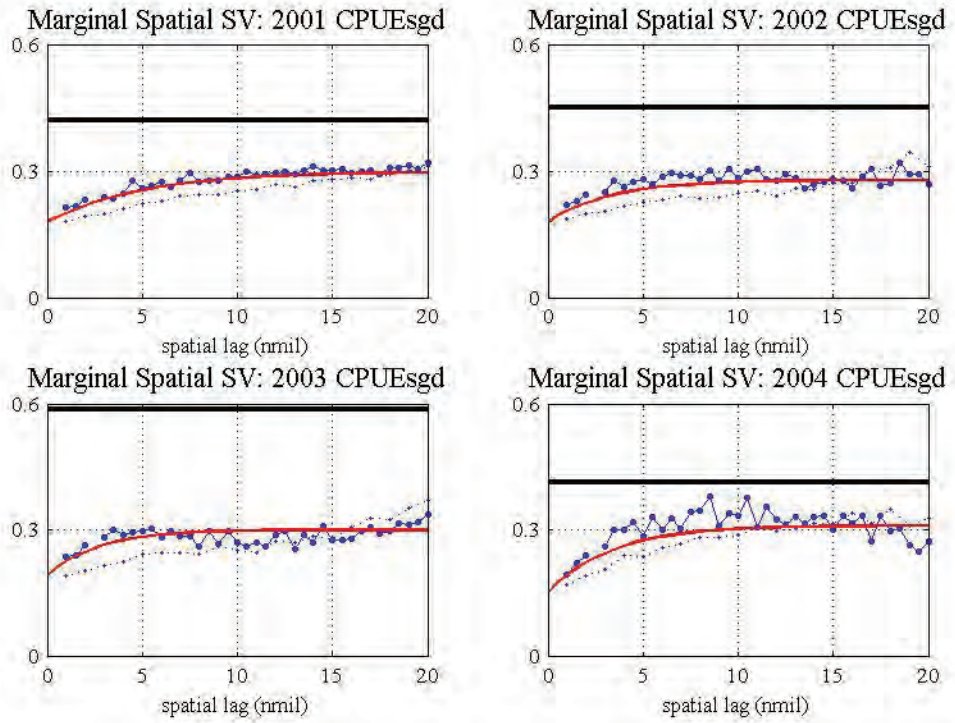


Figure 7.13. Marginal spatial semivariograms (blue) and fitted semivariogram models (red) of *CPUEsgd* for individual seasons 2001-2004. General relative semivariogram shown by dotted blue line. Variance of data set shown by solid black line.

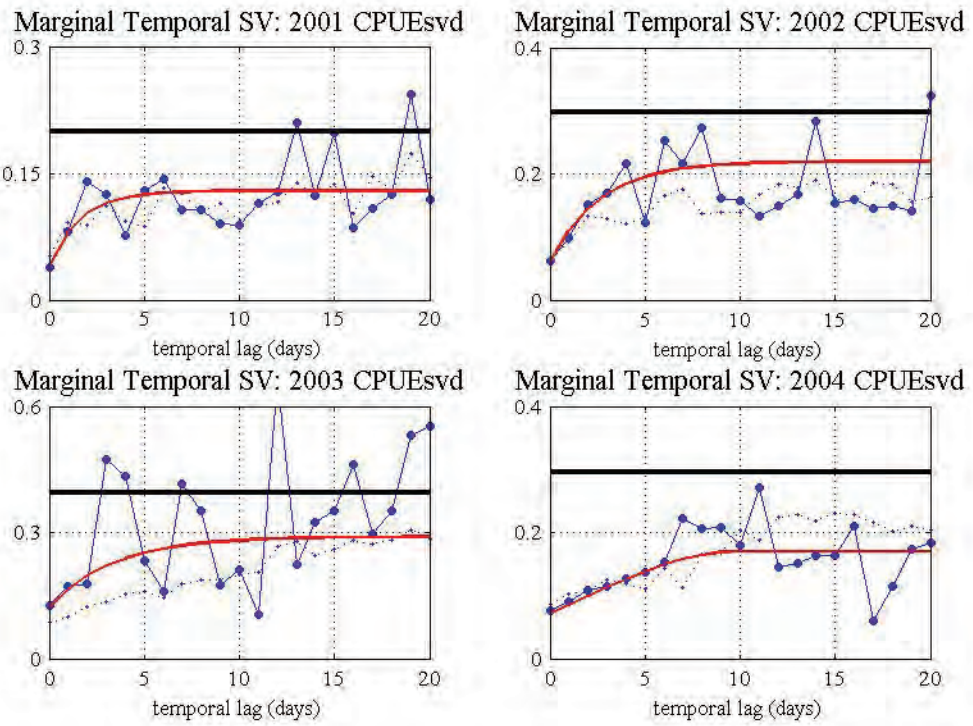


Figure 7.14. Marginal temporal semivariograms (blue) and fitted semivariogram models (red) of *CPUEsvd* for individual seasons 2001-2004. Pairwise relative semivariogram shown by dotted blue line. Variance of data set shown by solid black line.

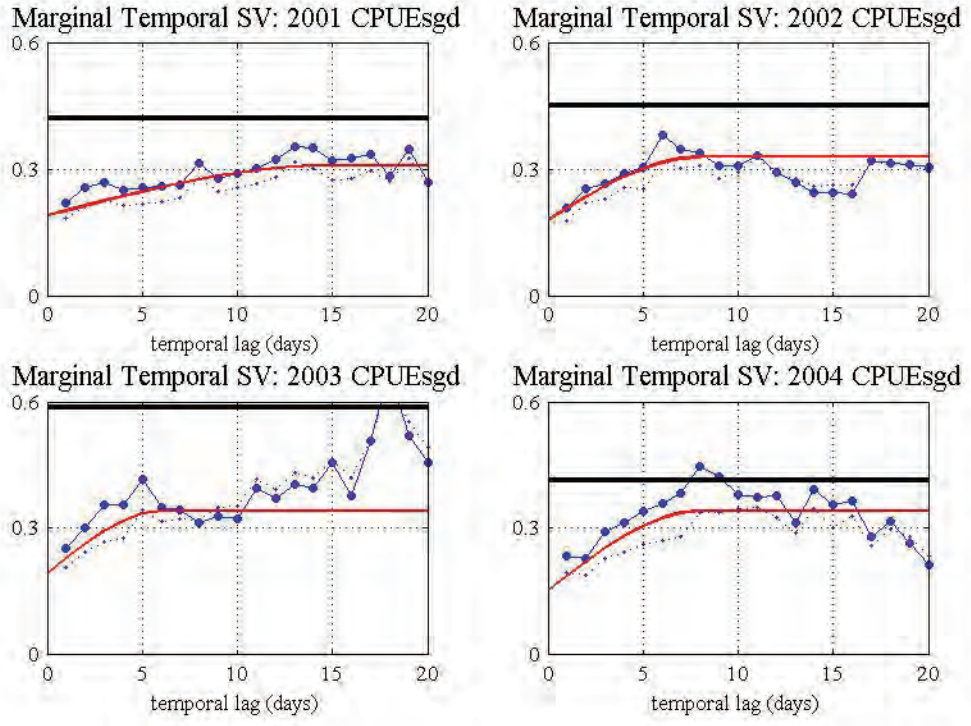


Figure 7.15. Marginal temporal semivariograms (blue) and fitted semivariogram models (red) of *CPUEsgd* for individual seasons 2001-2004. General relative semivariogram shown by dotted blue line. Variance of data set shown by solid black line.

Comparative plots of the fitted semivariogram models (Figure 7.16) for the marginal spatial and temporal semivariograms shows that the *CPUEsvd* marginal semivariograms have lower sills than those of the *CPUEsgd* data. This is to be expected as the *vd* data sets have comparatively lower variances. The shape of the marginal spatial variograms for the *CPUEsvd* data is somewhat similar for all seasons for the first 5 nmil but then the seasons show varying sills at which the variogram flattens out. The marginal temporal variogram models show similar ranges for the four seasons but varying sills. Season 2001 shows the lowest sill for both its marginal spatial and temporal semivariogram which is to be expected as it has the lowest data variance. The marginal variograms of the *CPUEsgd* data show even greater similarities across the four seasons, with similar short and long ranges for both spatial and temporal marginal variograms.

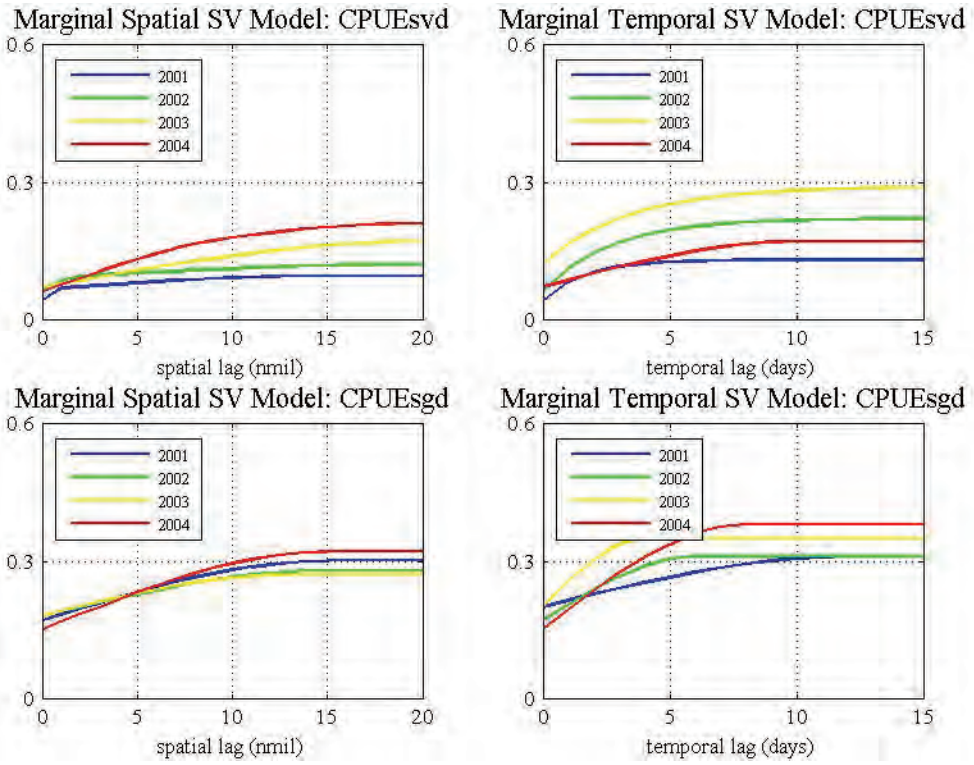


Figure 7.16. Marginal spatial and temporal semivariogram models of *CPUEsvd* and *CPUEsgd* for the individual seasons 2001-2004.

7.2.4 SPATIOTEMPORAL SEMIVARIOGRAM MODELLING

The product-sum semivariogram model was chosen to model the standardised catch rate semivariograms as it offered a flexible range of models that combined space-time variability, allowed varying ranges and sills in the spatial and temporal dimensions, as well as the overall sill. The product-sum model requires the modelling of the associated marginal spatial and temporal semivariograms, along with the global sill of the spatiotemporal variogram.

The marginal temporal and spatial experimental semivariograms seen in the previous section have high nuggets. Fitting a product-sum model with a high nugget effect for both the spatial and temporal marginal semivariogram models resulted in a spatiotemporal nugget effect that was considerably higher than the individual spatial and temporal nuggets leading to a poor fit of the overall model. Since the temporal and spatial nuggets of a given season are similar this can be interpreted as a global nugget. As the catch rate data involved spatially aggregated catch and effort values for each day (by vessel or grid) then the nugget effect could be considered primarily as a spatial nugget effect. In other words, the temporal nugget observed also included the effect of small spatial separations. This was particularly obvious for the *vd* data as their locations were not defined on a grid so the computation of the experimental marginal temporal semivariogram required the inclusion of a spatial separation at a small tolerance (less than 0.5 nmil) rather than the theoretical zero lag spacing in the spatial sense. In order to assure a good fit the marginal temporal variogram nugget and sill parameters were corrected to account for this deviation.

7.2.5 MODIFICATION OF MARGINAL TEMPORAL SEMIVARIOGRAM MODEL

If *tnug*, *tsill*, *snug*, *ssill* and *gsill* denote the parameters for the marginal semivariograms and the global sill respectively then the corrected marginal temporal variogram parameters, *tnug** and *tsill** are defined as

$$tsill^* = \frac{ssill - (snug - tsill)\dfrac{ssill}{snug} - gsill}{\dfrac{ssill}{snug} - 1}, \quad tnug^* = \frac{tnug - snug}{1 - \left(\dfrac{ssill + tsill^* - gsill}{ssill.tsill^*}\right)snug} \tag{7.1}$$

These modified parameters define the marginal temporal semivariogram model that produces a product-sum model, as defined by (2.59) and (2.60), which satisfies

$$\gamma_{st}(\mathbf{h}_s < 0.5, h_t > \mathrm{a}_t) = tsill \tag{7.2}$$

$$\gamma_{st}(\mathbf{h}_s < 0.5, h_t = 0) = tnug \,. \tag{7.3}$$

This ensures that the fitted product-sum model matches the experimental spatiotemporal semivariogram well at zero spatial separation (as fitted by the marginal temporal semivariogram model).

The corrected temporal nuggets for the *CPUEsvd* and *CPUEsgd* variables of the individual seasons of 2001 to 2004 are equal to, or close to, zero (Table 7.4). This modification can be regarded as an alternative method for implementing a global nugget effect by a single product-sum model. In the product-sum semivariogram model, the marginal spatial semivariogram model, and hence the spatial nugget effect, takes effect for all non-zero temporal lag spacings. This is the model assumed for the semivariograms of the prawn catch rates, with the marginal spatial semivariogram model containing the global nugget and the marginal temporal semivariogram modified to remove the effect of this global nugget.

Table 7.4. Modified marginal temporal semivariogram models of *CPUEsvd* and *CPUEsgd* for individual seasons 2001-2004.

Season	2001 vd	2001 gd	2002 vd	2002 gd	2003 vd	2003 gd	2004 vd	2004 gd
Nugget	0.00	0.00	0.00	0.00	0.06	0.00	0.01	0.00
1st structure								
Type	exp	sph	exp	sph	exp	sph	sph	sph
Range	5	15	8	8	10	6	10	12
Sill	0.09	0.13	0.20	0.19	0.19	0.17	0.13	0.28

7.2.6 PRODUCT-SUM SEMIVARIOGRAM MODELS

Product-sum models were constructed using the marginal spatial and (modified) temporal semivariogram models. Having described the spatial and temporal behaviour separately, the values of the semivariogram beyond the spatial and temporal ranges were examined to infer the global sill. The variance of the associated season was used as an indicator of the global sill but it is unreliable practice to force the global sill to equal the sample variance as they typically differ (Journel & Huigbregts, 1978; Goovaerts, 1997). The validity of the fitted product-sum models were checked for the values of the temporal, spatial and global sills using the diagnostic outlined in Section 2.2.8.4 to ensure the resulting space-time semivariogram function was negative definite. The product-sum models fitted using these corrected temporal variogram parameters supports the correction and facilitates the implementation of the product-sum model in the subsequent GSLIB kriging and simulation routines. A further modification of the GSLIB routines to allow for a distinct global nugget in combination with a product-sum model (or other spatiotemporal semivariogram models) is recommended for future work.

Product-sum semivariogram models (Figure 7.17) were fitted to the experimental semivariograms of *CPUEsvd* (Figure 7.1) using the marginal spatial and (modified) temporal semivariogram models (Table 7.5). The seasons 2001, 2002 and 2004 show great similarity in

their shape with slightly differing global sills. Season 2001 displays the smallest sill and season 2003 has the highest. The model for season 2003 displays different behaviour to the other seasons. For small spatial lags, the global sill is attained much quicker in the temporal direction. Comparison with the experimental semivariograms of Figure 7.1 indicates a good fit. The experimental semivariograms of *CPUEsgd* (Figure 7.2) were also fitted with product-sum semivariogram models (Figure 7.18). There is even more similarity across the seasons than for the *CPUEsvd* semivariogram models, with season 2003 displaying a slightly higher global sill. Comparison with the experimental semivariograms of Figure 7.2 indicates a good fit for the shape of the semivariograms. The allowable range of k to ensure permissibility of the models, given by (2.61), is satisfied for the *CPUEsvd* and *CPUEsgd* variables for all individual seasons (Table 7.5).

Table 7.5. Product-sum model parameters semivariograms of *CPUEsvd* and *CPUEsgd* for individual seasons 2001-2004.

Season	2001 vd	2001 gd	2002 vd	2002 gd	2003 vd	2003 gd	2004 vd	2004 gd
Spatial model								
Nugget	0.04	0.18	0.06	0.18	0.07	0.18	0.06	0.15
1st structure								
Type	sph	exp	sph	exp	sph	exp	sph	exp
Range	1	15	2	15	20	13	10	15
Sill	0.025	0.12	0.03	0.09	0.1	0.09	0.05	0.16
2nd structure								
Type	sph		sph				sph	
Range	15		20				20	
Sill	0.03		0.03				0.10	
Temporal model								
Nugget	0.00	0.00	0.00	0.00	0.06	0.00	0.01	0.00
1st structure								
Type	exp	sph	exp	sph	exp	sph	sph	sph
Range	5	15	8	8	10	6	10	12
Sill	0.09	0.13	0.20	0.19	0.19	0.17	0.13	0.28
Global sill	0.18	0.4	0.24	0.4	0.35	0.41	0.25	0.4
k	0.97	0.62	3.33	1.17	1.61	0.65	3.38	2.20
k range	(0,10.5]	(0,3.4]	(0,5.0]	(0,3.7]	(0,4.0]	(0,3.7]	(0,4.8]	(0,3.2]

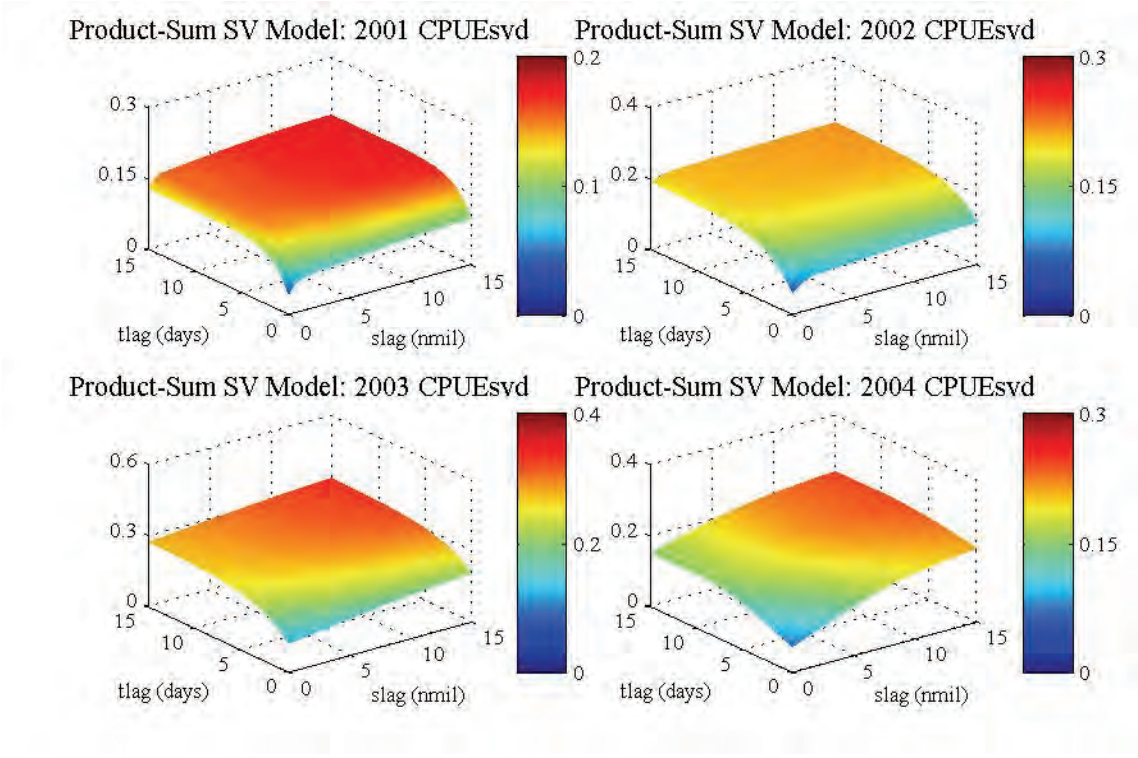


Figure 7.17. Product-sum semivariogram models for *CPUEsvd* of individual seasons 2001-2004 using semivariogram model parameters outlined in Table 7.5.

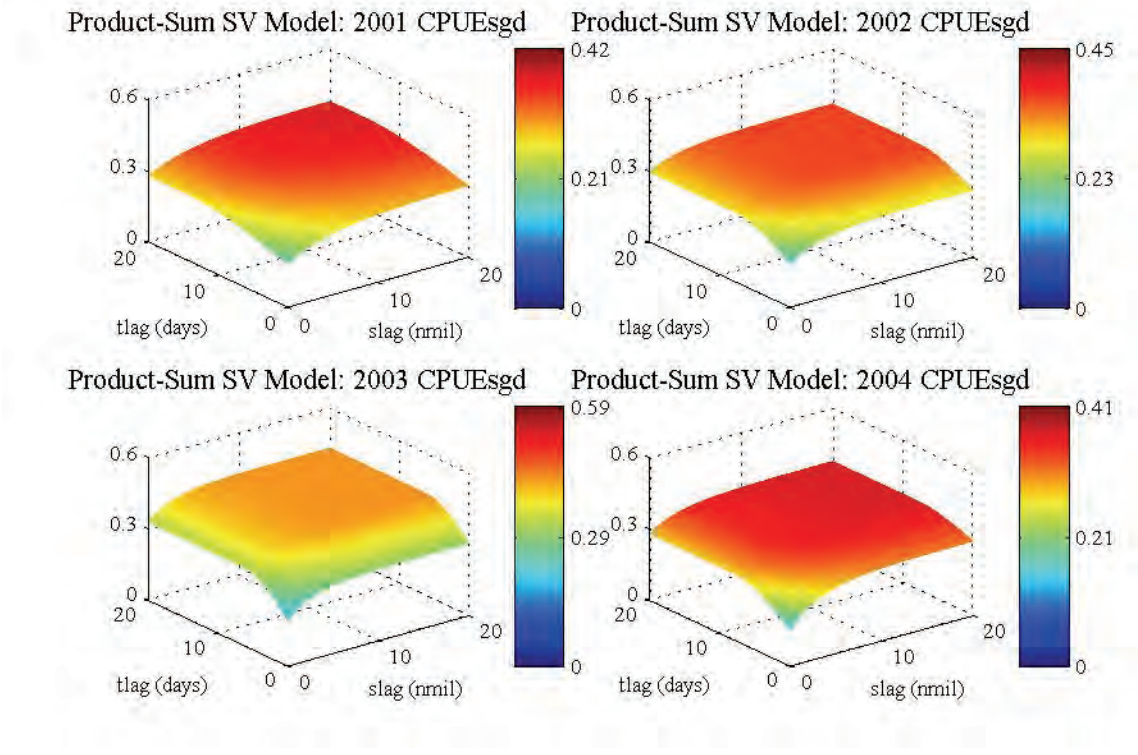


Figure 7.18. Product-sum semivariogram models for *CPUEsgd* of individual seasons 2001-2004 using semivariogram model parameters outlined in Table 7.5.

Although the product-sum model parameters were fitted manually to the experimental semivariograms, weighted sum of squares were considered for the fitted models. As for the marginal semivariogram models, temporal lags up to 15 days and spatial lags (0.5 nmil interval) up to 15 nmil were used in the calculations, and were weighted by the standardised number of pairs for each lag. Season 2001 and 2002 are the best performers for the *vd* and *gd* data respectively, with season 2003 showing the poorest fit for both. These varied results confirm the manual fitting of the semivariogram models. The objective of the semivariogram model is to capture the major spatiotemporal features of the catch rates, not just fit a model that is the closest possible to experimental values (Goovaerts, 1997). The models of spatiotemporal continuity are based on a view of the overall continuity evident in each season and the obvious similarities between each season confirming the presence of spatial and temporal continuity in the catch rate data.

Table 7.6. Weighted sum of squares (Equation 2.41) of product-sum semivariogram models of *CPUEsvd* and *CPUEsgd* for individual seasons 2001-2004.

Season	CPUEsvd	CPUEsgd
2001	0.105	0.047
2002	0.021	0.007
2003	0.144	0.017
2004	0.049	0.016

7.3 SPATIOTEMPORAL VARIOGRAPHY FOR COMBINED SEASONS

As the spatiotemporal semivariogram models for the individual seasons of 2001 through 2004 showed similarities, modelling was also undertaken for the combined seasons. As for the individual seasons, standardised catch rate data was obtained for the combined seasons by removing the temporal trend obtained by classical decomposition. The motivation in considering combined seasons was to produce a model of spatiotemporal continuity for the combined season 2001-2003 that could be used to model the spatiotemporal continuity of season 2004. This would allow estimation and simulation of the 2004 catch rates from the 2001-2003 catch rate data. The combined seasons considered included three seasons, 2001-2002, 2002-2003 and 2001-2003, not involving the 2004 season to be estimated. Additionally, three seasons involving 2004 were also considered: 2003-2004, 2002-2004 and 2002-2004. In reality, these seasons would not be available for a future season, but are considered here for investigative purposes. Spatiotemporal semivariograms were calculated using the GSLIB

gamvmod.exe program with only 1 directional (omnidirectional) semivariogram considered for the spatial plane. Directional semivariograms were not considered for the combined seasons as the individual seasons showed no evidence of anisotropy. Moreover, the directional and omnidirectional semivariograms were similar for spatial separation distances less than 15 nmil and temporal separations less than 15 days.

7.3.1 EXPERIMENTAL SPATIOTEMPORAL SEMIVARIOGRAMS

The experimental spatiotemporal semivariogram for the 2001-2003 combined season showed similar behaviour to those of the individual seasons with increasing value for increasing temporal and spatial lag spacings (Figure 7.19). The experimental spatiotemporal semivariograms for the other combined seasons considered also showed similar behaviour to those of the individual seasons, with the combined season 2003-2004 having less structure than the other combined seasons (APPENDIX I). The variances of the combined data sets (Table 7.7) are in a similar range to those of the individual seasons (0.202 to 0.362) suggesting the sills reached by the variograms will be similar.

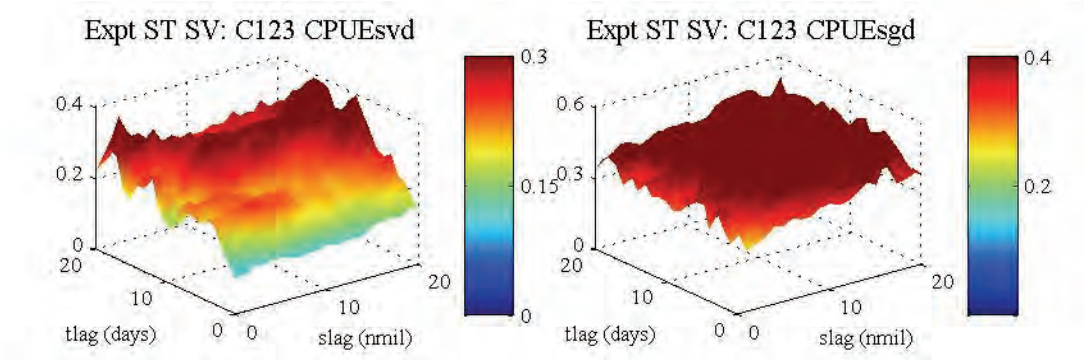


Figure 7.19. Experimental spatiotemporal semivariograms of *CPUEsvd* and *CPUEsgd* for combined season 2001-2003.

Table 7.7. *CPUEsvd* and *CPUEsgd* data variance for combined seasons of 2001-2004.

Combined Season	Variance *CPUEsvd*	Variance *CPUEsgd*
C12	0.257	0.437
C23	0.352	0.516
C123	0.306	0.482
C34	0.356	0.503
C234	0.355	0.485
C1234	0.316	0.467

7.3.2 PRODUCT-SUM SEMIVARIOGRAM MODELS

The marginal spatial and temporal variograms (APPENDIX I) were similar across the six combined seasons and showed similar structure to that seen in the marginal semivariograms of

the individual seasons. The spatial and temporal ranges of the combined seasons are comparable to those seen in the variograms of the individual seasons, with spatial continuity levelling off around 20 nmil and temporal correlation diminishing around 10 days.

The number of pairs associated with the marginal spatial and temporal variograms showed peaks in the early lag spacings as seen for the individual seasons (APPENDIX I).With increased number of pairs at all temporal and lag spacings, the marginal semivariograms showed smoother patterns of continuity than seen for the individual seasons. The similarity of the continuity measures across the combined seasons for both the *vd* and *gd* data was supportive of the use of a model of spatiotemporal continuity for the (standardised) catch rates.

Semivariogram models, for the combined season 2001-2003, fitted to the marginal spatial variograms and temporal variograms (Figure 7.20) of the combined seasons used similar models to the associated individual seasons (Figure 7.16). The other combined seasons showed similar model fits for their marginal semivar iograms (APPENDIX I). The *CPUESvd* variograms involved a nugget and two spherical structures and the *CPUEsgd* variograms involved a nugget and two exponential structures (Table 7.8) which are similar to those of the individual seasons (Table 7.2).

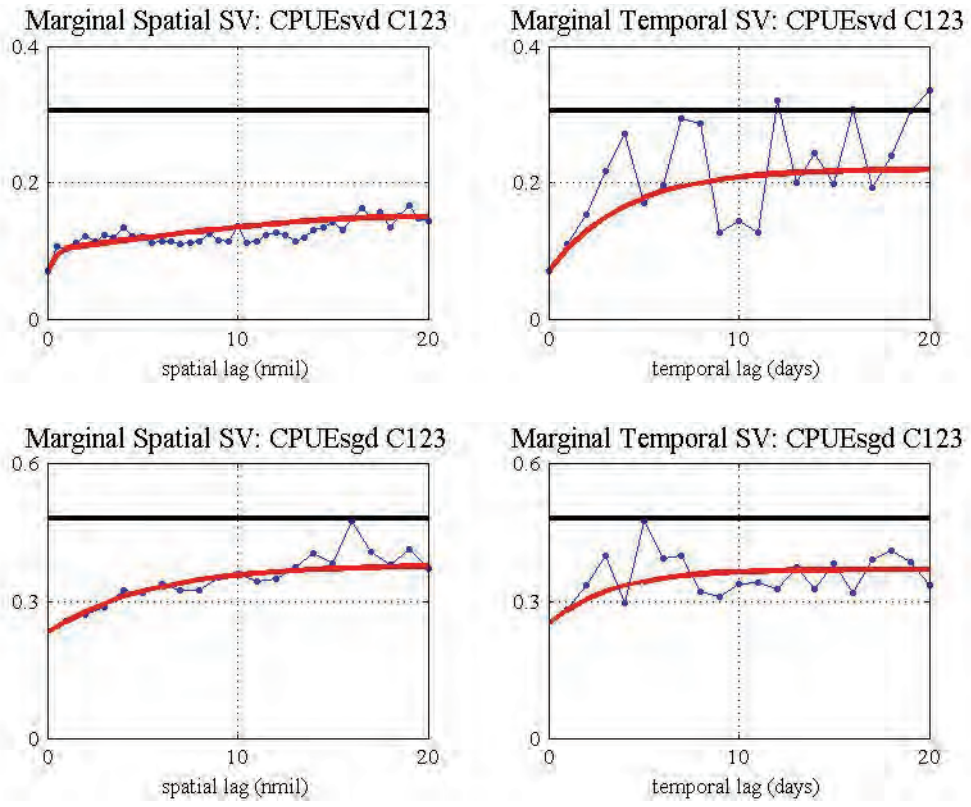


Figure 7.20. Marginal spatial (left) and temporal (right) semivariograms (blue) and fitted semivariogram models (red) of *CPUEsvd* (top row) and *CPUEsgd* (bottom row) for combined season 2001-2003. Variance of data set shown by solid black line.

Table 7.8. Marginal spatial and temporal semivariogram model parameters of *CPUEsvd* and *CPUEsgd* for combined seasons of 2001-2004.

Season	C12		C23		C123		C34		C234		C1234	
	vd	gd	vd	gd	vd	gd	vd	gd	vd	gd	vd	gd
Spatial model												
Nugget	0.05	0.22	0.08	0.22	0.07	0.23	0.11	0.22	0.08	0.21	0.07	0.22
1st structure												
Type	sph	sph	sph	sph	sph	exp	sph	exp	sph	exp	sph	exp
Range	2	12	1	12	1	16	20	20	1	20	1	16
Sill	0.03	0.15	0.03	0.15	0.03	0.15	0.09	0.15	0.02	0.16	0.03	0.15
2nd structure												
Type	sph		sph		sph				sph		sph	
Range	20		20		20				20		20	
Sill	0.05		0.05		0.05				0.07		0.06	
Temporal model												
Nugget	0.05	0.22	0.08	0.22	0.07	0.25	0.11	0.3	0.08	0.25	0.08	0.24
1st structure												
Type	exp	sph	exp	sph	exp	exp	exp	exp	exp	exp	exp	exp
Range	12	10	15	10	12	10	15	10	14	10	15	10
Sill	0.14	0.12	0.2	0.12	0.15	0.12	0.17	0.1	0.18	0.12	0.17	0.11

As for the individual seasons, the combined seasons showed high spatial and temporal nuggets, and therefore the marginal temporal semivariogram model parameters were adjusted to ensure a good fit for the product-sum models to the experimental spatiotemporal semivariograms. Product-sum semivariogram models computed for the combined seasons using the parameters of the marginal spatial and (modified) temporal semivariogram models are very similar across the combined seasons (APPENDIX I.5). The global sill of each model was fitted just under the variance of the associated (combined) season. The model for the 2001-2003 combined season (Figure 7.21) is similar to the product-sum models seen previously for the individual seasons 2001 to 2004 (Figure 7.17 and Figure 7.18).

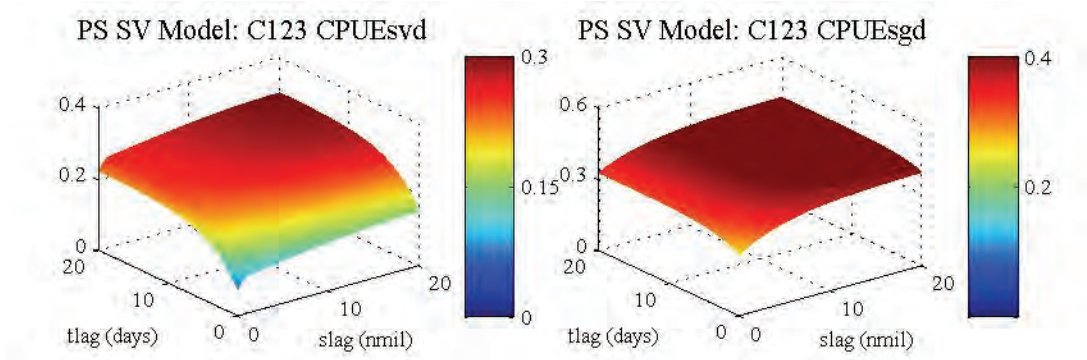


Figure 7.21. Product-sum semivariogram models, combined seasons, 2001-2003.

Table 7.9. Product-sum model parameters semivariograms of *CPUEsvd* and *CPUEsgd* for combined seasons.*Minimum *k* is 0.

Season	C12		C23		C123		C34		C234		C1234	
	vd	gd	vd	gd	vd	gd	vd	gd	vd	gd	vd	gd
Spatial model												
Nugget	0.05	0.22	0.08	0.22	0.07	0.23	0.11	0.22	0.08	0.21	0.07	0.22
1st structure												
Type	sph	sph	sph	sph	sph	exp	sph	exp	sph	exp	sph	exp
Range	2	12	1	12	1	16	20	20	1	20	1	16
Sill	0.03	0.15	0.03	0.15	0.03	0.15	0.09	0.15	0.02	0.16	0.03	0.15
2nd structure												
Type	sph		sph		sph				sph		sph	
Range	20		20		20				20		20	
Sill	0.05		0.05		0.05				0.07		0.06	
Temporal model												
Nugget	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.13	0.00	0.06	0.01	0.03
1st structure												
Type	exp	sph	exp	sph	exp	exp	exp	exp	exp	exp	exp	exp
Range	12	10	15	10	12	10	15	10	14	10	15	10
Sill	0.15	0.22	0.21	0.13	0.15	0.20	0.19	0.16	0.19	0.17	0.20	0.16
Global sill	0.25	0.42	0.35	0.48	0.30	0.46	0.35	0.48	0.34	0.48	0.3	0.46
k	1.64	2.10	0.60	0.50	0.08	1.72	1.14	1.65	0.59	1.39	2.11	1.41
k max*	6.56	2.70	4.76	2.70	6.63	2.63	5.00	2.70	5.29	2.70	4.74	2.70

7.4 SPATIOTEMPORAL VARIOGRAPHY OF NORMAL SCORES

Semivariogram models were also required for the normal scores of the standardised catch rate for use in a Gaussian simulation process. Once normal scores were obtained (*CPUEsN*), the semivariogram estimation and modelling process was conducted as for the *CPUEs* data using the GSLIB *gamvmod.exe* program. Experimental semivariograms for the four individual seasons were computed (Figure 7.22 and Figure 7.23).

Marginal spatial and temporal variograms for the normal scores of the *vd* and *gd* standardised catch rates were computed and modelled (Figure 7.24-Figure 7.27). The marginal semivariogram models showed more consistency across seasons for the data aggregated by grid, and considerably higher relative nuggets as seen for the standardised catch rates.

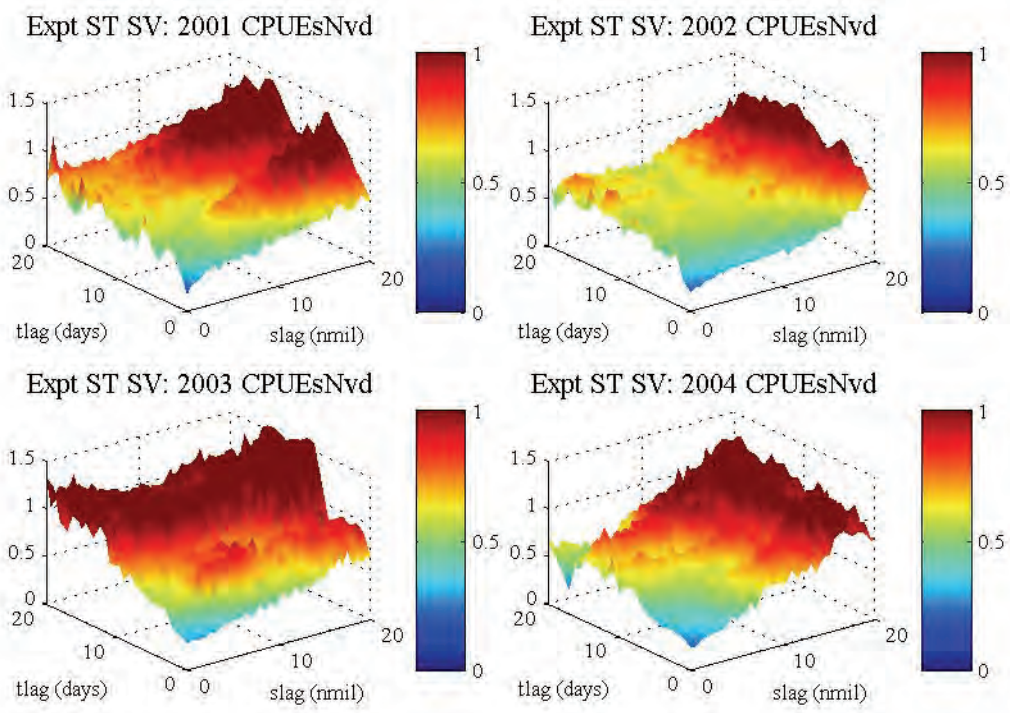


Figure 7.22. Experimental spatiotemporal semivariograms of *CPUEsNvd* for seasons 2001-2004.

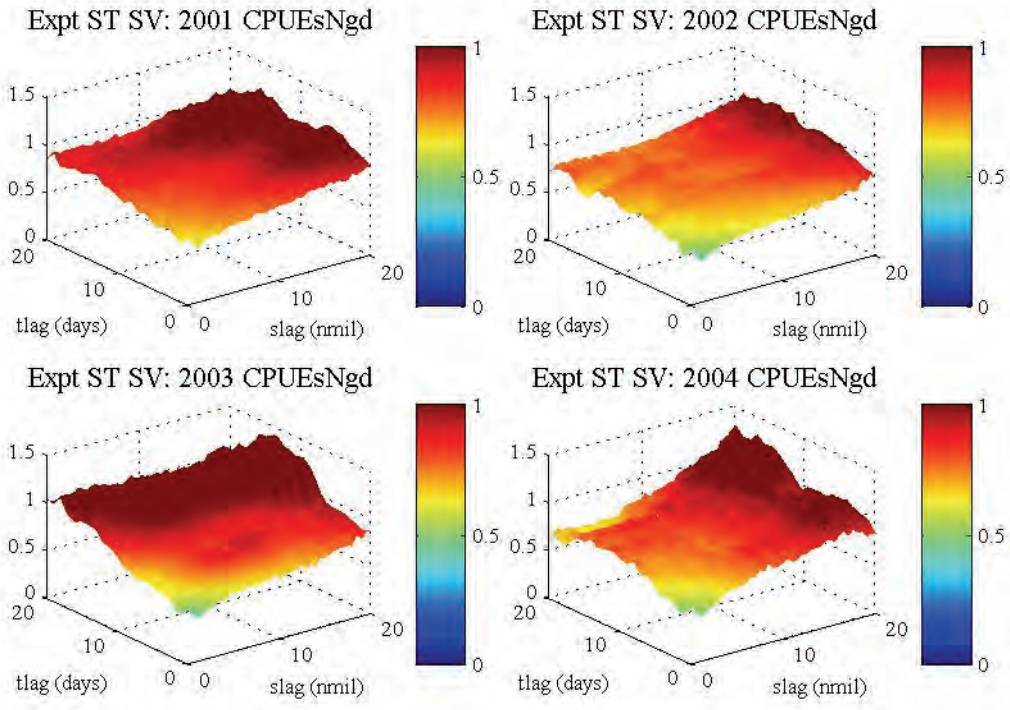


Figure 7.23. Experimental spatiotemporal semivariograms of *CPUEsNgd* for seasons 2001-2004.

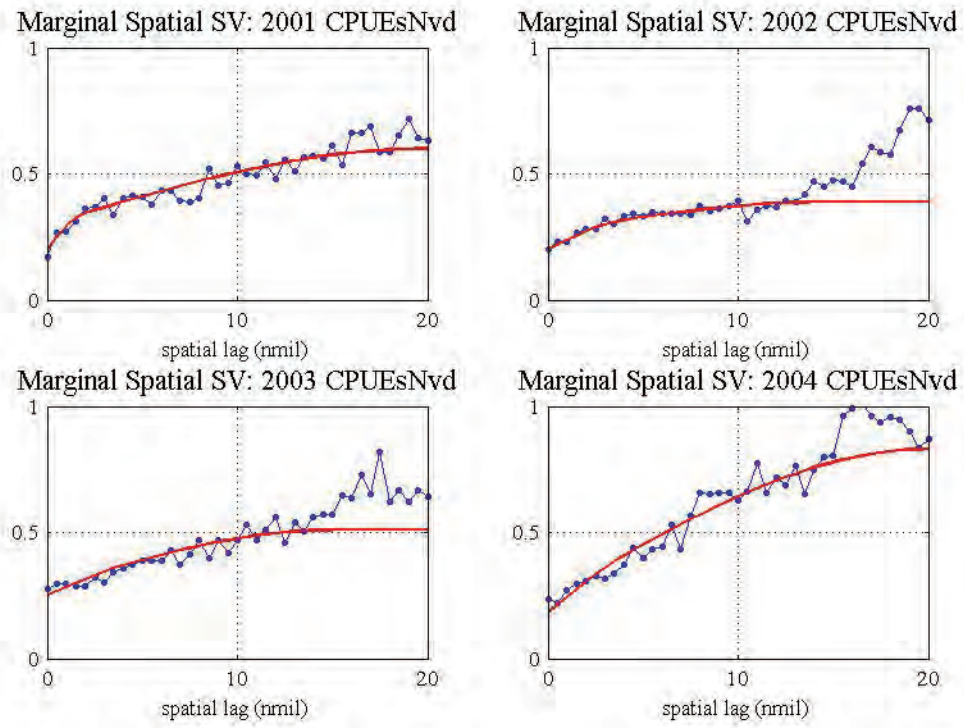


Figure 7.24. Marginal spatial semivariograms (blue) and fitted semivariogram models (red) of *CPUEsNvd* for individual seasons 2001-2004.

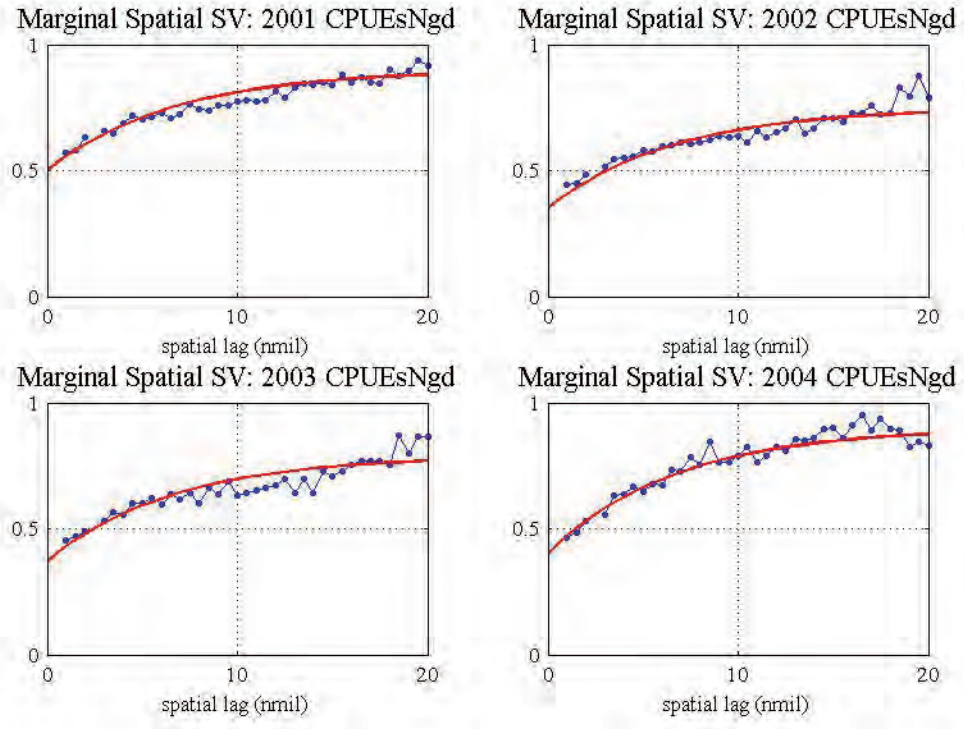


Figure 7.25. Marginal spatial semivariograms (blue) and fitted semivariogram models (red) of *CPUEsNgd* for individual seasons 2001-2004.

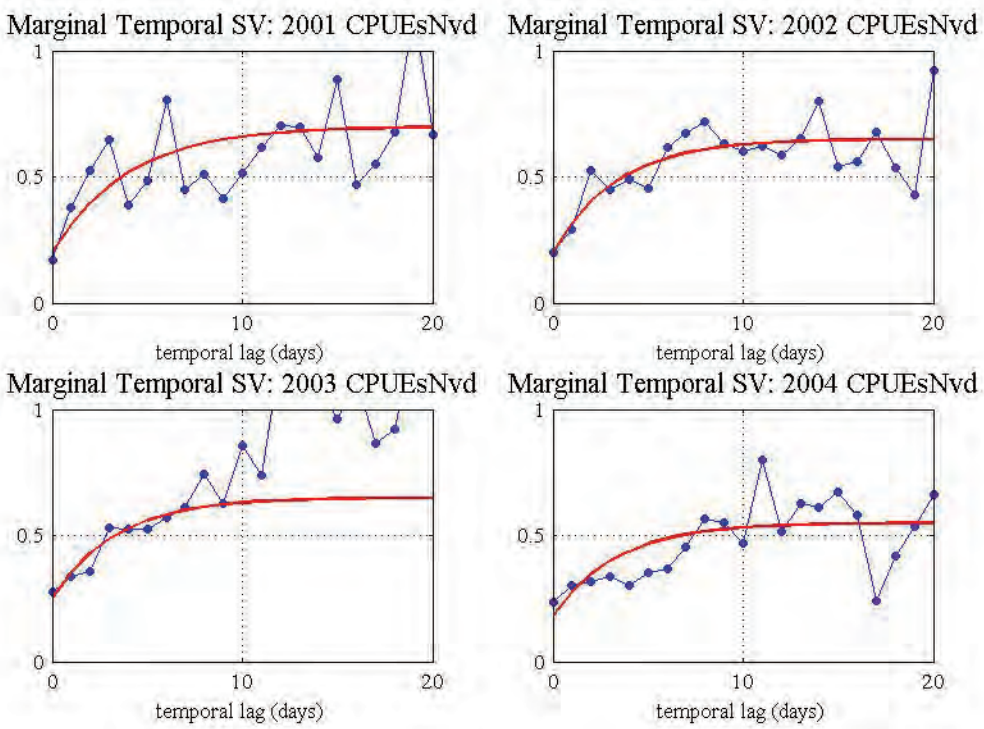


Figure 7.26. Marginal temporal semivariograms (blue) and fitted semivariogram models (red of *CPUEsNvd* for individual seasons 2001-2004.

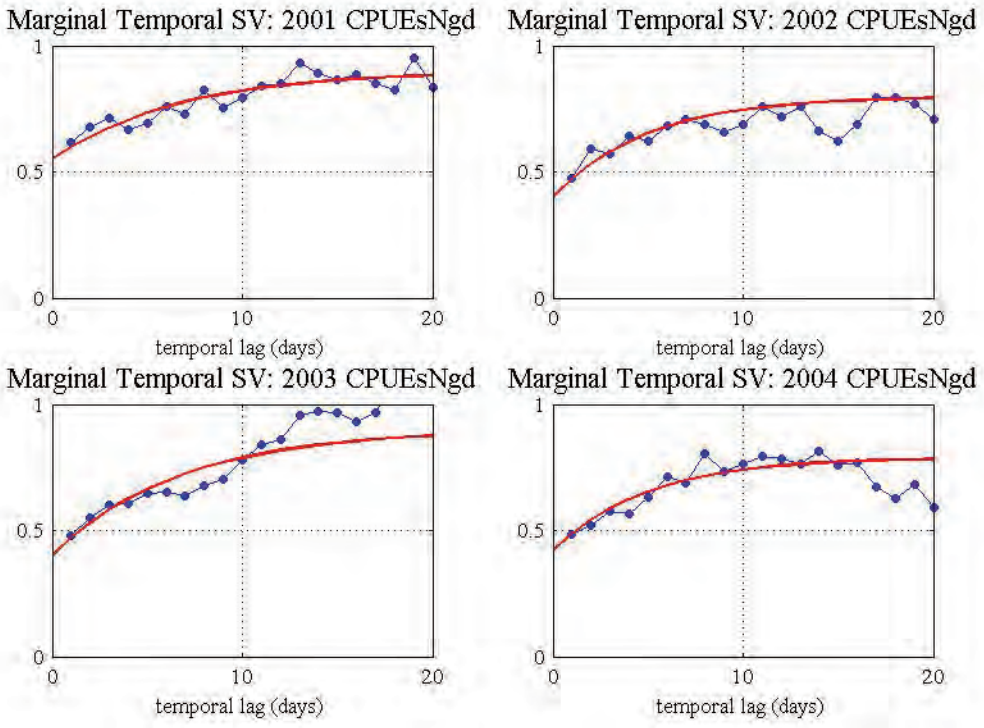


Figure 7.27. Marginal temporal semivariograms (blue) and fitted semivariogram models (red) of *CPUEsNgd* for individual seasons 2001-2004.

As for the standardised catch rates, the marginal temporal semivariogram model parameters were adjusted to account for the global nugget effect, and were then used, along with the marginal spatial semivariogram models and global sill, to construct product-sum

semivariogram models (Table 7.10, Figure 7.28 and Figure 7.29) to describe the spatiotemporal continuity.

Table 7.10. Product-sum model parameters semivariograms of *CPUEsNvd* and *CPUEsNgd* for individual seasons 2001-2004.

Season	2001 vd	2001 gd	2002 vd	2002 gd	2003 vd	2003 gd	2004 vd	2004 gd
Spatial model								
Nugget	0.2	0.5	0.2	0.35	0.25	0.37	0.18	0.4
1st structure								
Type	sph	exp	sph	exp	sph	exp	sph	exp
Range	2	20	4	20	4	20	4	20
Sill	0.1	0.4	0.07	0.4	0.03	0.42	0.05	0.5
2nd structure								
Type	sph		sph		sph		sph	
Range	20		15		15		20	
Sill	0.3		0.3		0.23		0.6	
Temporal model								
Nugget	0.00	0.10	0.00	0.07	0.00	0.05	0.00	0.03
1st structure								
Type	exp	sph	exp	sph	exp	sph	sph	sph
Range	5	15	8	8	10	6	10	12
Sill	0.55	0.68	0.46	0.56	0.51	0.77	0.43	0.59
Global sill	1	1	1	1	1	1	1	1
k	0.45	0.97	0.12	0.80	0.08	0.94	0.72	0.93
k range	(0,1.7]	(0,1.1]	(0,1.8]	(0,1.3]	(0,2]	(0,1.2]	(0,1.2]	(0,1.1]

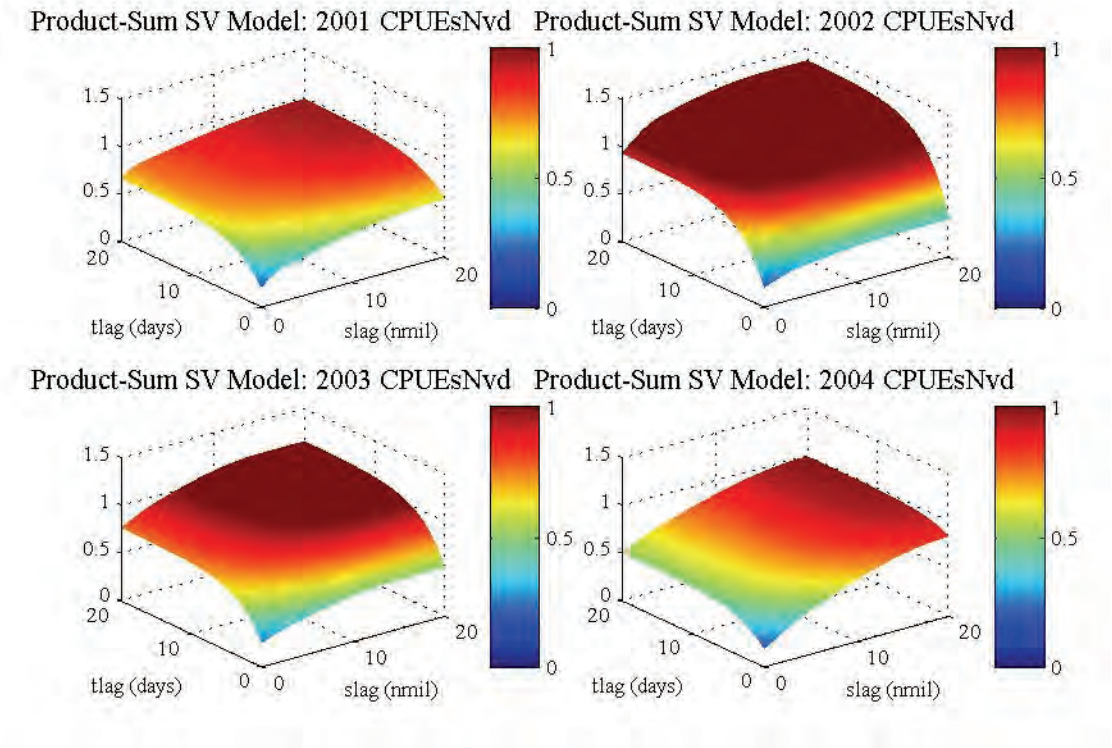


Figure 7.28. Product-sum semivariogram models of *CPUEsNvd* for individual seasons 2001-2004.

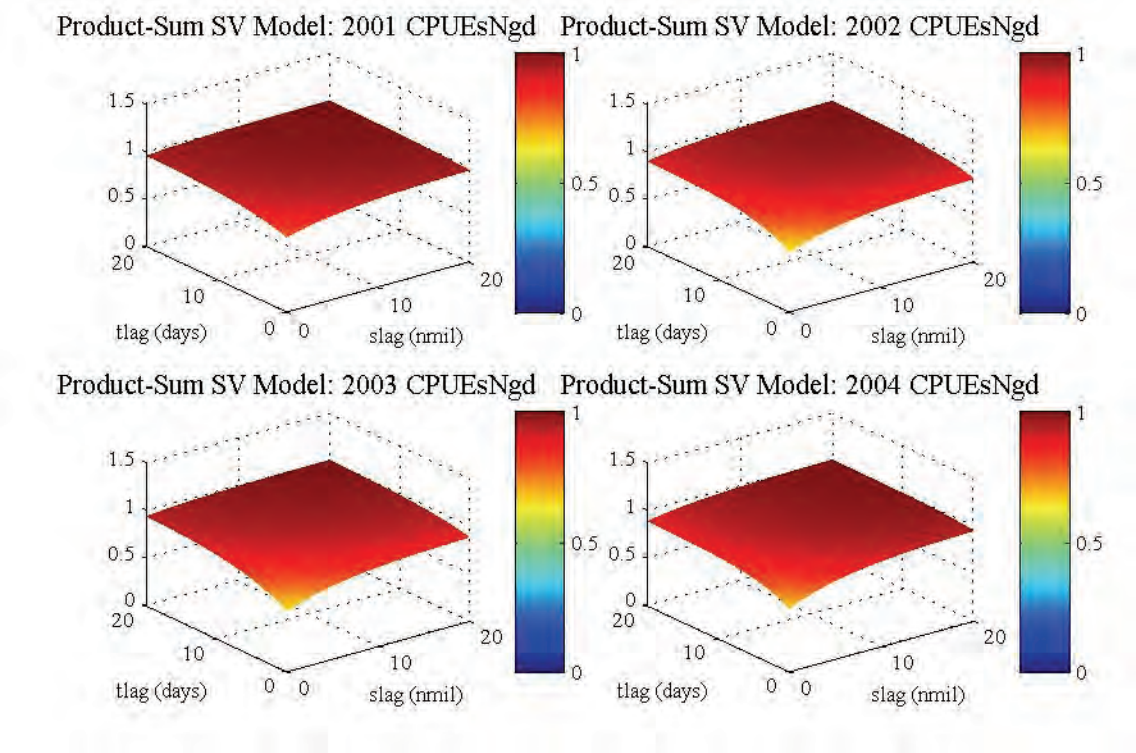


Figure 7.29. Product-sum semivariogram models of *CPUEsNgd* for individual seasons 2001-2004.

The semivariograms, and their fitted models, of the normal scores of the individual seasons 2001-2003 showed strong similarities. The evidence of similarity within these seasons supported the development of a model of spatiotemporal continuity for the combined seasons 2001-2003 to estimate the spatiotemporal continuity of season 2004, as conducted in the modelling process of the standardised catch rates. This model would facilitate the geostatistical simulation of the catch rates of a subsequent (lunar) month of 2004. Recall that the simulation process required normally distributed data, and therefore a model of spatiotemporal continuity for the normal scores of the standardised catch rates. The product-sum semivariogram model obtained for the combined seasons 2001-2003 (Table 7.11, Figure 7.30 and Figure 7.31) displayed similar behaviour to those of the individual seasons 2001-2003.

Table 7.11. Product-sum semivariogram model parameters of *CPUEsNvd* for combined season 2001-2003.

Marginal Spatial Model			Marginal Temporal model		
Nugget		0.20	Nugget		0.00
1st structure	Type	spherical	1st structure	Type	exponential
	Range	3		Range	10
	Sill	0.04		Sill	0.52
2nd structure	Type	spherical			
	Range	15		Global Sill	1.00
	Sill	0.3		k (max k)	0.22(1.9)

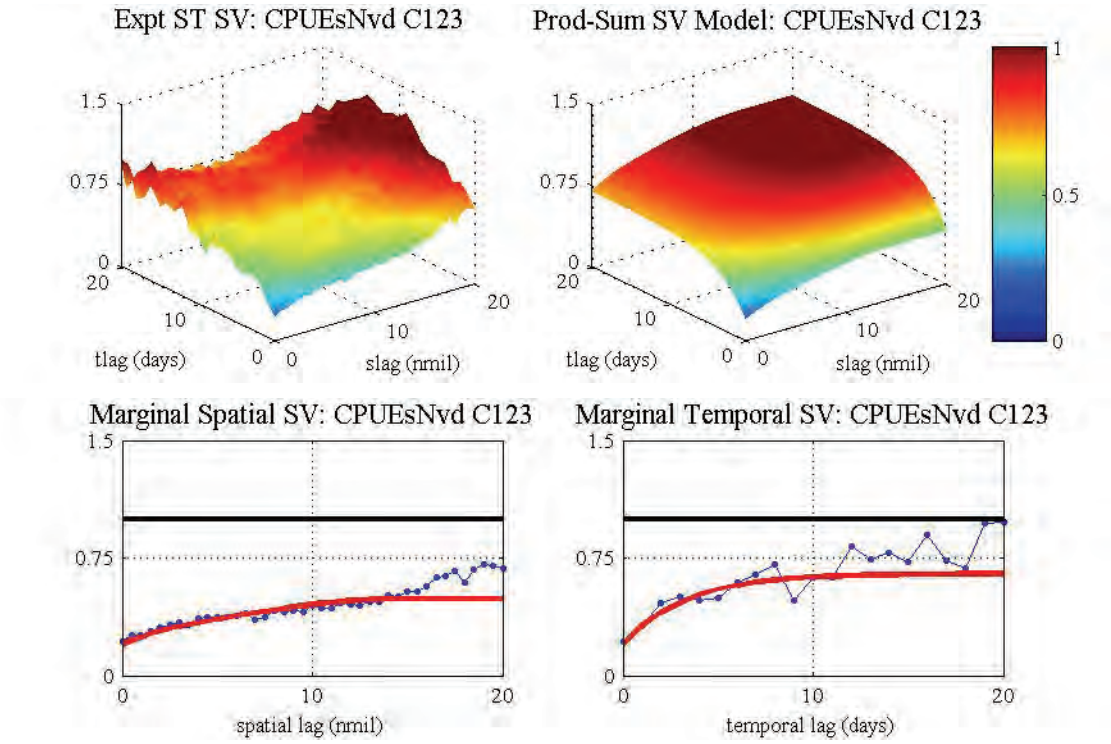


Figure 7.30. Experimental semivariogram (top left), product-sum model (top right), marginal spatial semivariogram (bottom left) and (unmodified) marginal temporal semivariogram (bottom right) of *CPUEsvdN* for combined season 2001-2003.

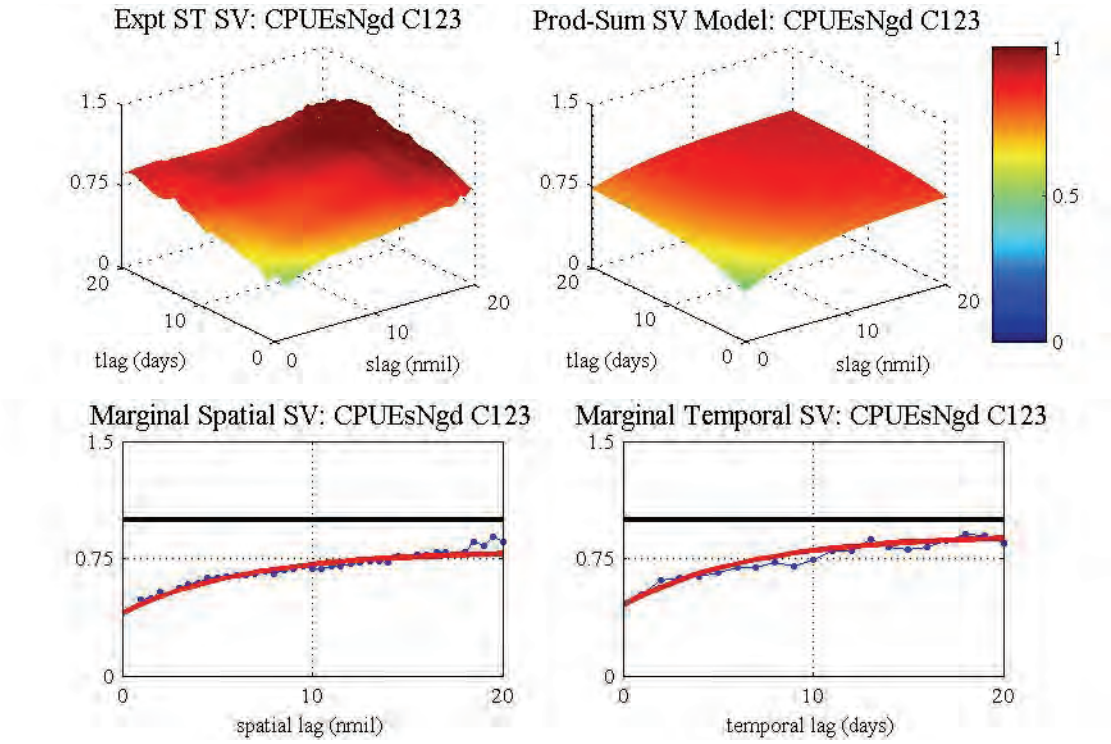


Figure 7.31. Experimental semivariogram (top left), product-sum model (top right), marginal spatial semivariogram (bottom left) and (unmodified) marginal temporal semivariogram (bottom right) of *CPUEsgdN* for combined season 2001-2003.

7.5 SELECTION OF SEMIVARIOGRAM MODEL

In order to estimate or simulate *CPUE* values for the April to October lunar months of season 2004, the spatiotemporal models of continuity obtained from the standardised catch rate data and normal scores of the seasons 2001 to 2003 was required to estimate the spatiotemporal model of continuity for the 2004 season. Ordinary spatiotemporal kriging of each lunar month of the 2004 standardised catch rate data can be conducted using the GSLIB program *kt3dnew.exe* along with the product-sum semivariogram model of combined seasons 2001 to 2003 and the data of the previous lunar month of 2004. In a similar manner, sequential Gaussian simulation of the catch rate normal scores can be conducted using the semivariogram model of the 2001-2003 normal scores. The *CPUEs* estimates or simulations are then multiplied by the classical decomposition trend model of the combined season 2001-2003 to obtain estimates or simulations of the catch rates. The actual *CPUE* data of each lunar month of season 2004 will be used in this study to evaluate the model. In reality, the locations to be estimated/simulated in a future season are unknown, so estimates are made on a grid over the entire region in order to highlight areas of high of low catch rates, along with their uncertainty.

Thus far, the spatiotemporal continuity of two catch rate variables (*CPUEsvd* and *CPUEsgd*) has been modelled. Since the models of spatiotemporal continuity of the two variables showed many similarities, it was decided to only proceed with estimation using one model. As outlined previously, the model for the combined season 2001-2003 was chosen as a (estimated) measure of the spatiotemporal continuity of the 2004 season based on its similarities with all three individual seasons involved. Moreover these three seasons occur directly before 2004 and with no obvious changes in management for the 2004 season there is no reason to expect a significant change in the continuity of the catch rates. The question then arose whether to use the product-sum variogram model of the *CPUEsvd* or *CPUEsgd* of the combined season 2001-2003 (Figure 7.21). The *CPUEsvd* model was chosen as the variograms of the *CPUEsvd* data for all seasons show longer ranges and lower nugget effects in both the spatial and temporal direction for all individual and combined seasons, indicative of a stronger measure of spatiotemporal continuity. This was supported by the associated use of the *vd* data obtained by aggregating by vessel and removed the problems inherent in the data aggregated by grid, which combined data aggregation methods when shot data was considered for aggregation by grid with data already aggregated by vessel.

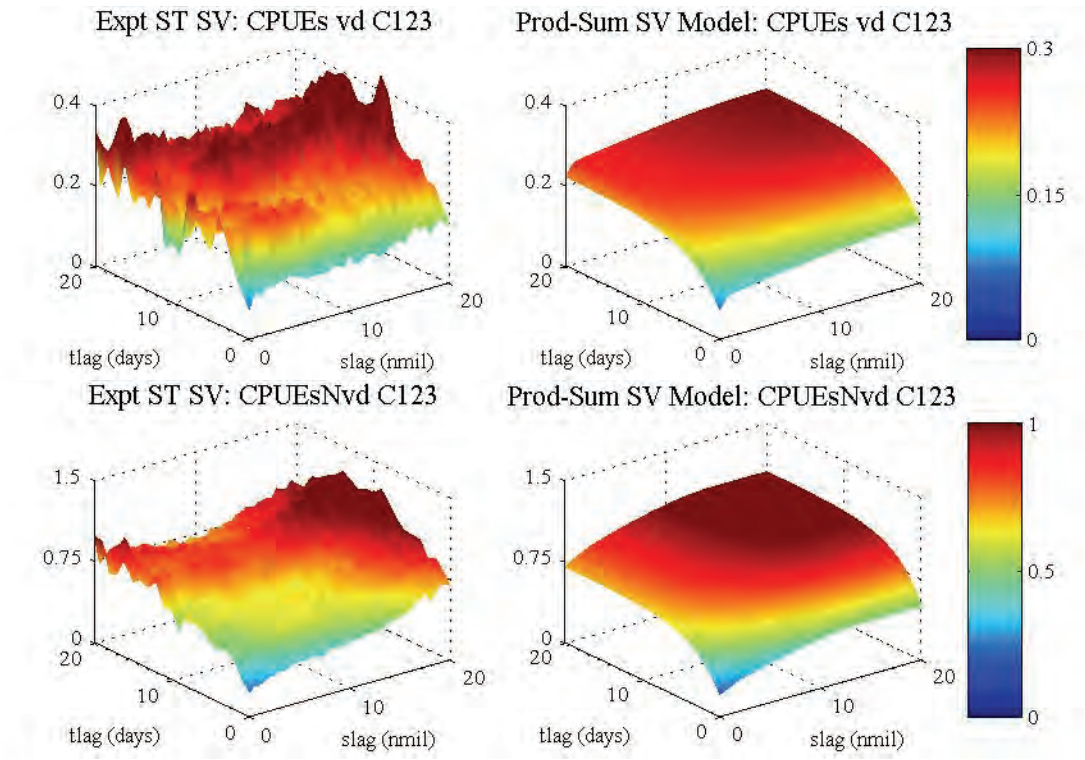


Figure 7.32. Experimental semivariogram (left) and product-sum models (right) of *CPUEsvd* (top row) and *CPUEsNvd* (bottom row) for combined season 2001-2003.

7.6 CHAPTER SUMMARY

The spatiotemporal continuity of the standardised catch rate variable was modelled by considering the spatiotemporal semivariograms of the *CPUEsvd* and *CPUEsgd* variables. Both variables showed evidence of spatial and temporal continuity for the individual and combined seasons. This pattern of continuity was modelled by a product-sum semivariogram model, which was constructed from the marginal spatial and temporal variograms along with the global sill exhibited by the experimental semivariogram. Relative semivariograms were used to infer the parameters of each model.

A modification of the marginal temporal variogram parameters (nugget and sill) was developed to ensure a good fit to the experimental spatiotemporal semivariograms. This modification presents an alternative method for the implementation of a global nugget effect in a spatiotemporal semivariogram model within a single product-sum model.

The individual seasons of 2001-2003 showed similar patterns of spatiotemporal continuity so it was reasonable to assume that the 2004 season would display a similar pattern. This meant that a model of the spatiotemporal continuity for season 2004 could be estimated from the behaviour of the seasons 2001-2003. Rather than using a model of a single season or averaging the parameters of the three individual models, a model of spatiotemporal continuity

was derived for the combined season 2001-2003. This allowed the pooling of data over the three individual seasons. The model for the combined season of 2001-2003 showed similar patterns of continuity to those seen in the models for the individual seasons. This was deemed the more appropriate model to use as an estimate of the spatiotemporal continuity for the individual season 2004 that could be used to allow geostatistical estimation and simulation of the 2004 catch rate values. The semivariogram models for the standardised catch rates of the *vd* data and the associated normal scores were selected for use in a subsequent estimation and simulation process for the 2004 catch rates as they displayed the strongest measure of spatiotemporal continuity.

PART III.

GEOSTATISTICAL SPATIOTEMPORAL ESTIMATION AND SIMULATION

CHAPTER 8

SPATIOTEMPORAL ESTIMATION

8.1 INTRODUCTION

In this chapter we consider estimation of the king prawn catch rate for season 2004 using the product-sum semivariogram model obtained in Chapter 7 for the standardised king prawn catch rate of the combined season 2001-2003. The model aimed to predict the catch rates of each month of the 2004 catch rates using the previous month's data as conditioning data. With the electronic collection of data to become more prevalent in Western Australia fisheries, faster data validation and processing will be possible. Therefore, this forecasting method could be a viable tool to assist fishery scientists in the monitoring of their fishery.

The suitability of the model was first assessed for the conditioning data of 2004. The cross-validation procedure was used to assess the suitability and predictive ability of the model of spatiotemporal continuity for estimation of the standardised catch rates of season 2004. If the model could capture the spatiotemporal variation of the catch rates in a lunar month of 2004, then it would be reasonable to use that semivariogram model to predict catch rates for the next lunar month. The catch rate data of 2004 were transformed to standardised catch rates using the global trend model of the combined season 2001-2003 and the cross validation process conducted on the lunar monthly standardised catch rate data sets. Cross validation estimates of the 2004 standardised catch rates were back-transformed to catch rate values using the trend model.

Estimates of standardised catch rates were then made at the spatiotemporal locations (day and spatial location) of the 2004 data using only the 2004 data up to but not including the prediction month to assess the forecasting ability of the model. As for the cross-validation procedure, estimates were transformed to catch rate values using the global trend model. A similar forecasting process was used to compute grid estimates of the catch rates across the fishing region for the 2004 season. Daily estimates were obtained over the grid and used to create (lunar) weekly averages. Estimates at the data locations and grid estimates compared favourably with the actual data values (Denham & Mueller, 2009). This paper was presented at the Geostats2008 conference (APPENDIX A). Forecasts were also made incorporating additional survey data to help infer catch rate estimates in locations just opened for fishing

(Denham & Mueller, 2010) with results presented at the GeoENV VII conference (APPENDIX A).

8.2 KRIGING PROCESS

The existence of a model of spatiotemporal dependence makes it possible to estimate the values of a variable at unsampled locations. Kriging algorithms are a method for deriving least-squares linear regression estimates by using the information of nearby data related to the variable being estimated. The weight assigned to each datum in the estimation process depends on the kriging model chosen, which in turn relates to the model for the local mean. In this study we use ordinary kriging (OK) that considers the local mean to be unknown but stationary. Spatiotemporal effects are included via reference to the autocorrelation structure of the data set, as summarized by the fitted covariance model.

The kriging process can be used to conduct cross-validation of the estimation process with a particular spatiotemporal model of continuity. Cross-validation is the process of removing one datum at a time from the data set and re-estimating this value from the remaining data using the semivariogram model. It can be repeated to compare different semivariogram models or different search parameters. Interpolated and actual values are compared, and the model/search parameters that yield the most accurate predictions is retained. If the cross-validation process validates the use of a kriging process with its associated semivariogram model and search parameters then subsequent estimation can be performed at unsampled locations.

8.3 CROSS VALIDATION OF THE SPATIOTEMPORAL SEMIVARIOGRAM MODEL

Cross-validation was performed to assess how well the semivariogram model of the combined season 2001-2003 captured the spatiotemporal variation of the 2004 standardised king prawn catch rate data. Cross validation was conducted independently on each monthly data set of the 2004 season by estimating the standardised king prawn catch rate at each spatiotemporal sample point of that month in turn after excluding the sample value at that spatiotemporal location. The modified GSLIB program *kt3dnew.exe* (De Cesare et al., 2002) was used along with the associated parameter file (APPENDIX G.3).

Kriging was performed using the standardised king prawn catch rate data for the month of 2004 being estimated and the spatiotemporal semivariogram model for the combined 2001-2003 season (Figure 8.1). After several trials of various values, a minimum of 4 sample data and

a maximum of 20 were required to enable estimation. This ensured a relatively consistent number of points used for estimation across the region and did not demand excessive computation time. Eight lunar months (March-October) of 2004 involving three weeks (L,N,Q) in each were estimated via a cross-validation process. Estimates are not made for the week surrounding the full moon in all months as the fishing region is typically closed during these periods.

Table 8.1. Product-sum semivariogram model parameters of *CPUEsvd* for combined season 2001-2003.

Marginal Spatial Model		Marginal Temporal model	
Nugget	0.07	Nugget	0.00
1st structure Type	spherical	1st structure Type	exponential
Range	1	Range	12
Sill	0.03	Sill	0.15
2nd structure Type	spherical		
Range	20	Global Sill	0.30
Sill	0.05		

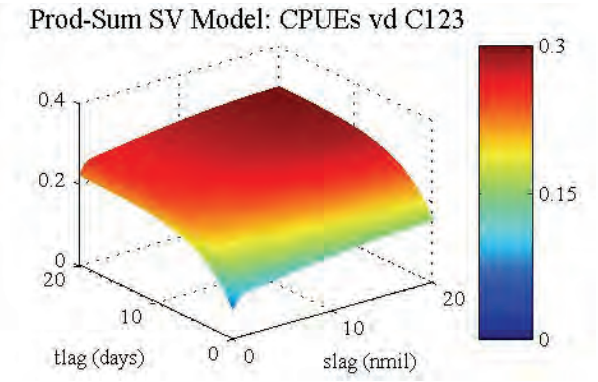


Figure 8.1. Product-sum model of *CPUEsvd* for the combined season 2001-2003.

Identification of the closest 20 data in the search procedure of the GSLIB *kt3dnew* program required the definition of a space-time distance metric by converting absolute measures of spatial and temporal separation (i.e. nmil and days, respectively) into relative measures based on their proportion of the maximum spatial and temporal search radii. The search radii have a dual purpose as they also determine the maximum spatial and temporal separation of points to be used in the estimation process. Cross-validation of each month was conducted using four search radii combinations to assess their impact (Table 8.2). Moreover, the impact of estimating using spatial-only or spatial and temporal information could also be assessed. The first search neighbourhood involved only the spatially closest 20 points from the same day as the point being estimated, but limited to a 5 nmil radius. A long range (20 nmil)

spatial neighbourhood was also used to enable the closest points up to 20 nmil away, which effectively meant 20 points would always be used. Two spatiotemporal neighbourhoods were considered to weight the relative importance of spatial and temporal information. The first neighbourhood allowed both spatial and temporal information whilst the second ensured a significant amount of temporal information was included as a temporal separation of 20 days was considered "as close" as a spatial separation of 5 nmil.

Table 8.2. Spatial and temporal search neighbourhoods.

	Spatial Search Radius	Temporal Search Radius
Spatial only – short range	5 nmil	0 days
Spatial only – long range	20 nmil	0 days
Spatial and temporal (1)	20 nmil	20 days
Spatial and temporal (2)	5 nmil	20 days

Using spatial-only information (spatial radius 5 nmil, temporal radius 0 days) to produce cross-validation estimates for a given month allowed only data located on the same day to be used to produce the weighted estimate at each location. The estimates produced showed relatively small and consistent mean errors, along with reasonable correlations with the actual 2004 data across the first five months (Figure 8.2). For the last three months the errors and correlations, were more variable which was partially due to the reduced number of estimates made. Increasing the spatial search radii to 20 nmil enabled more estimates to be made but this was achieved at the expense of greater errors and lower correlations with the actual values.

Cross-validation estimates computed using the spatiotemporal search neighbourhoods included the addition of data separated temporally (and spatially) from the location being estimated, along with the spatially (but not temporally) separated information. The best results were obtained using 20 nmil and 20 days respectively. The spatial radius of 20 nmil was set large to give the desired ratio for temporal information. In weeks where the spatial-only estimates produced reasonable estimates (small errors/good correlation with actual values), the additional temporal information allowed more estimates with slightly smaller errors (Figure 8.2) and slightly increased correlation with true values (Figure 8.3). In some cases, the effect was negligible. However, in weeks where the spatial-only estimates were unsatisfactory, the inclusion of temporal information improved the performance of the estimation with larger proportion of estimated locations (Figure 8.3, top row), smaller errors (Figure 8.2) and increased correlation with true values (Figure 8.3, bottom row),. It appears that in the absence of closely

related spatial information, relevant temporal information can improve the performance of the cross-validation estimates.

The inclusion of temporal information affects the number of spatial-only data included as the overall amount of data included does not change (unless specified). An alternative search neighbourhood has been implemented in another program (Spadavecchia, 2008) to allow for the specification of exact number of data included from each temporal separation rather than utilising a pseudo-distance metric. This is recommended for future work.

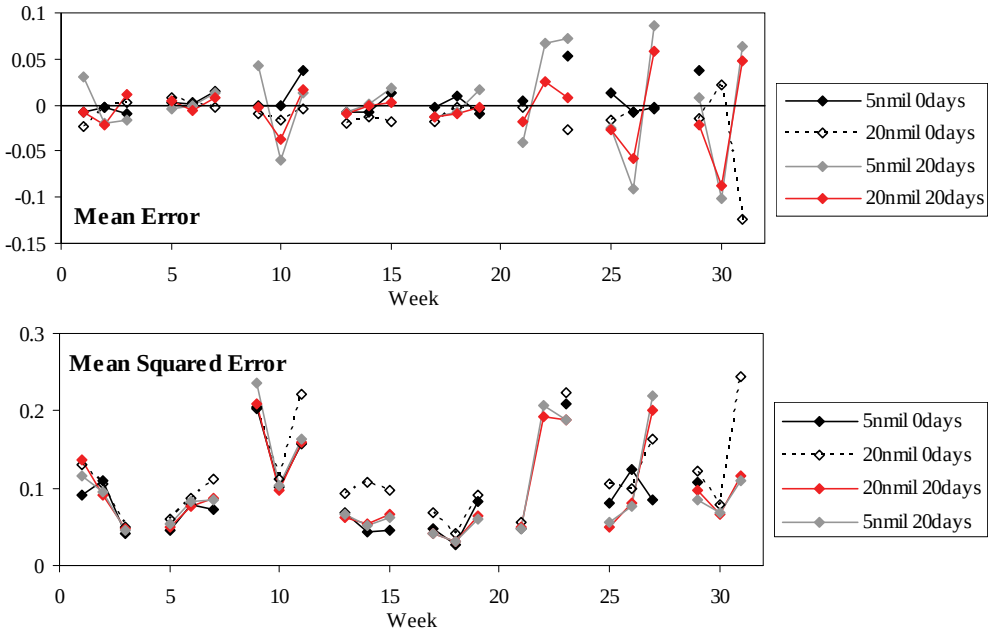


Figure 8.2. Mean error (top) and mean squared error (bottom) of cross validation results for the lunar weeks of 2004 using the four search neighbourhoods outlined in Table 8.2.

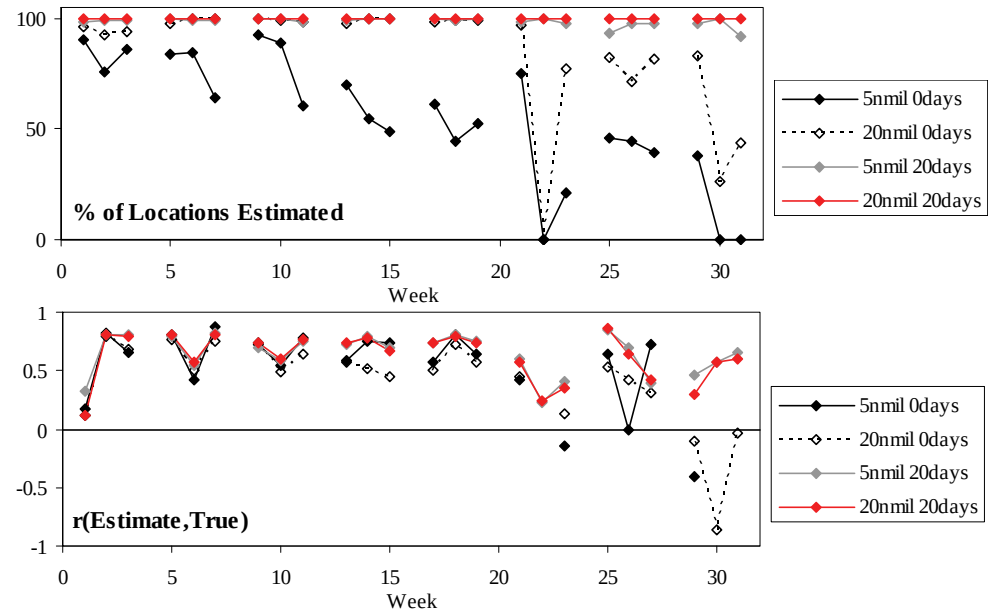


Figure 8.3. Percentage of locations estimated (top) and correlation of estimates with actual catch rate values (bottom) of cross validation results for the lunar weeks of 2004 using the four search neighbourhoods outlined in Table 8.2.

Although eight (lunar) months were estimated, results will be shown in this section for the months of May and July, with the results for all other months contained in APPENDIX J. The weeks of May were chosen as one example as the ENA opens at the beginning of the month resulting in relatively high catch rates. The weeks of July were chosen as a second example as they are indicative of the second half of the fishing season.

The standardised king prawn *CPUE* values have been reproduced over the months of May and July (Figure 8.4 and Figure 8.5). This is indicative of the estimates across the 2004 season (APPENDIX J). Across the month of May, there are regions of high and low values which have been reproduced well by the estimates, as have the mainly low values evident in July. All sample spatiotemporal locations in May and July have been estimated by the cross validation process (Table 8.3).

Across the season of 2004 only two sample locations did not produce cross validation estimates, one in March and one in October (APPENDIX J). The mean and median of the cross validation estimates are similar to those of the actual standardised catch rate values whilst the variance of the estimates is lower than that of the corresponding sample values for all weeks. The smoothing inherent in the kriging process results in the minimum/maximum of the estimates being higher/lower than the corresponding sample values.

Table 8.3. Summary statistics of *CPUEs* cross validation estimates for the lunar weeks of May and July 2004.

	MayL		MayN		MayQ		JulL		JulN		JulQ	
	Sample	CV Est	Sample	CV Est	Sample	CV Est	Sample	CV Est	Sample	CV Est	Sample	CV Est
Mean	1.652	1.627	1.228	1.265	0.917	0.911	0.694	0.704	0.700	0.707	0.659	0.654
Med	1.745	1.746	1.213	1.232	0.844	0.810	0.677	0.723	0.684	0.721	0.564	0.573
Var	0.455	0.221	0.152	0.065	0.382	0.219	0.091	0.055	0.086	0.060	0.136	0.068
Skew	-0.006	-0.477	0.112	-0.086	1.285	0.927	0.043	-0.385	0.316	0.045	0.818	0.562
Min	0.207	0.577	0.188	0.390	0.036	0.222	0.038	0.164	0.150	0.244	0.063	0.237
Max	3.527	2.557	2.357	1.796	3.271	2.243	1.459	1.095	1.495	1.248	1.668	1.321
N	196	196	203	203	165	165	197	197	202	202	166	166

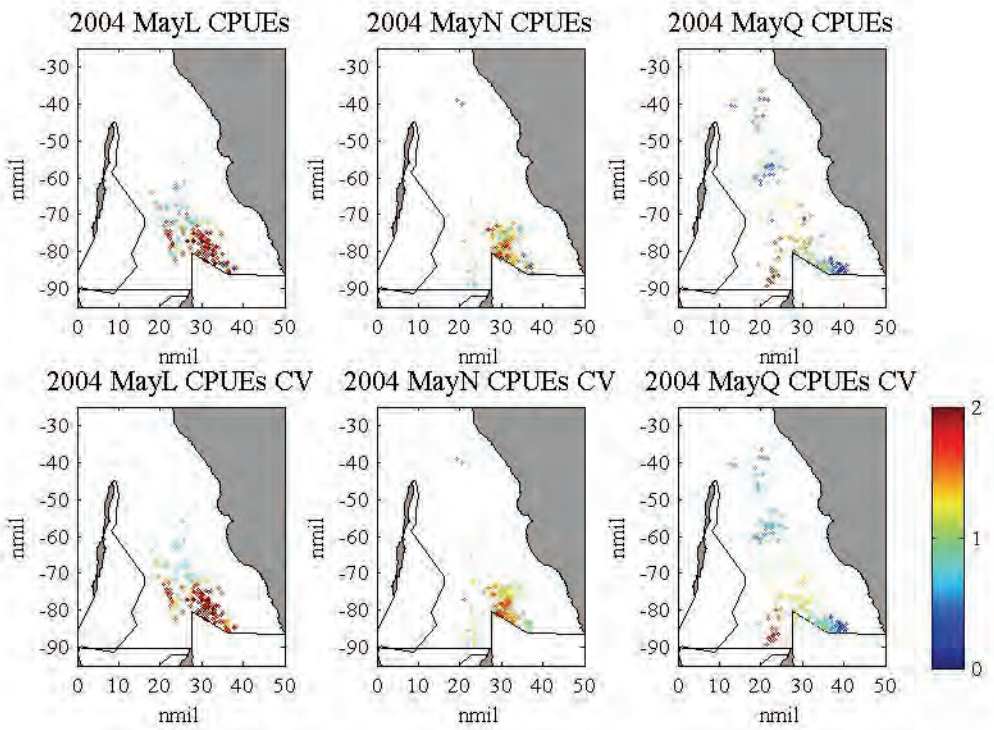


Figure 8.4. *CPUEsvd* values (top row) and cross validation estimates (bottom row) for lunar weeks of May 2004.

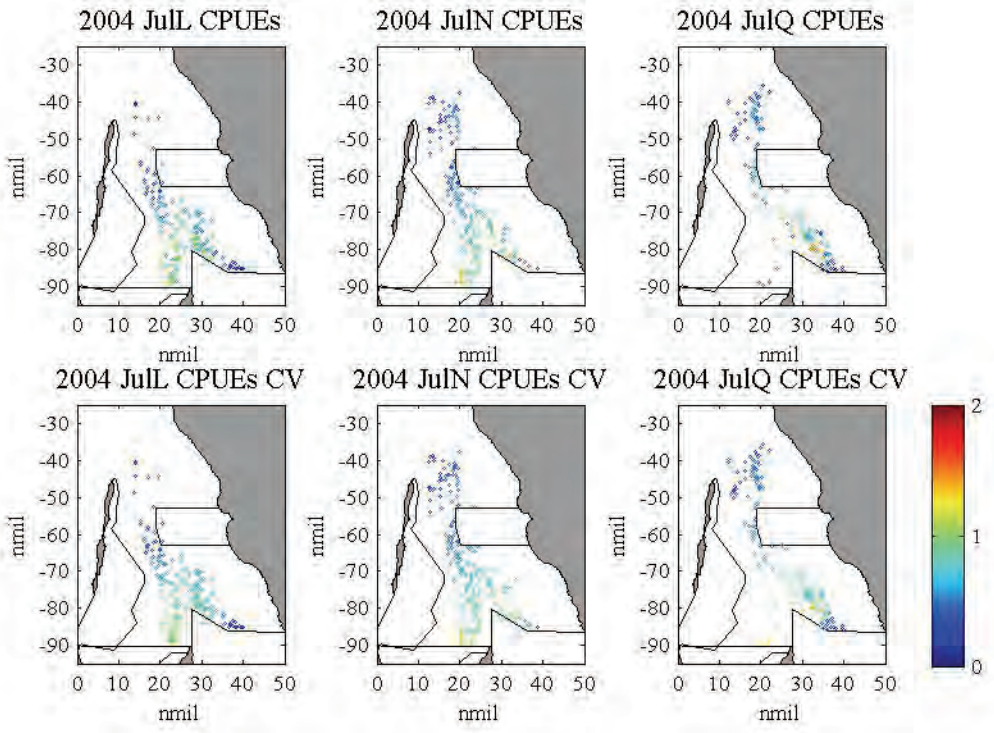


Figure 8.5. *CPUEsvd* values (top row) and cross validation estimates (bottom row) for lunar weeks of July 2004.

The range of cross-validation error values of May and July (top row, Figure 8.6 and Figure 8.7) are spread across the Shark Bay region with no obvious pattern of high or low error values. Scatterplots of the sample values against the estimated values (second row) show clouds close to the first bisector with correlation coefficients varying between 0.6 and 0.8 for May and

July indicating strong positive correlation between the estimates and sample values. Histograms of the standardised estimation errors (third row) indicate an approximately normal distribution. Scatterplots of the estimates against their standardised estimation errors (bottom row) show clouds close centred about the zero error line and display no obvious structural bias. The July clouds show less spread than those of May. The correlation coefficients vary between -0.09 and 0.07 for May and July indicating little or no correlation between the estimates and their errors. There a few outliers in May but none in July, with outliers being defined as those outside the 99% confidence limit of a normal distribution. Cross validation estimates for the other months of 2004 show similar behaviour (APPENDIX J).

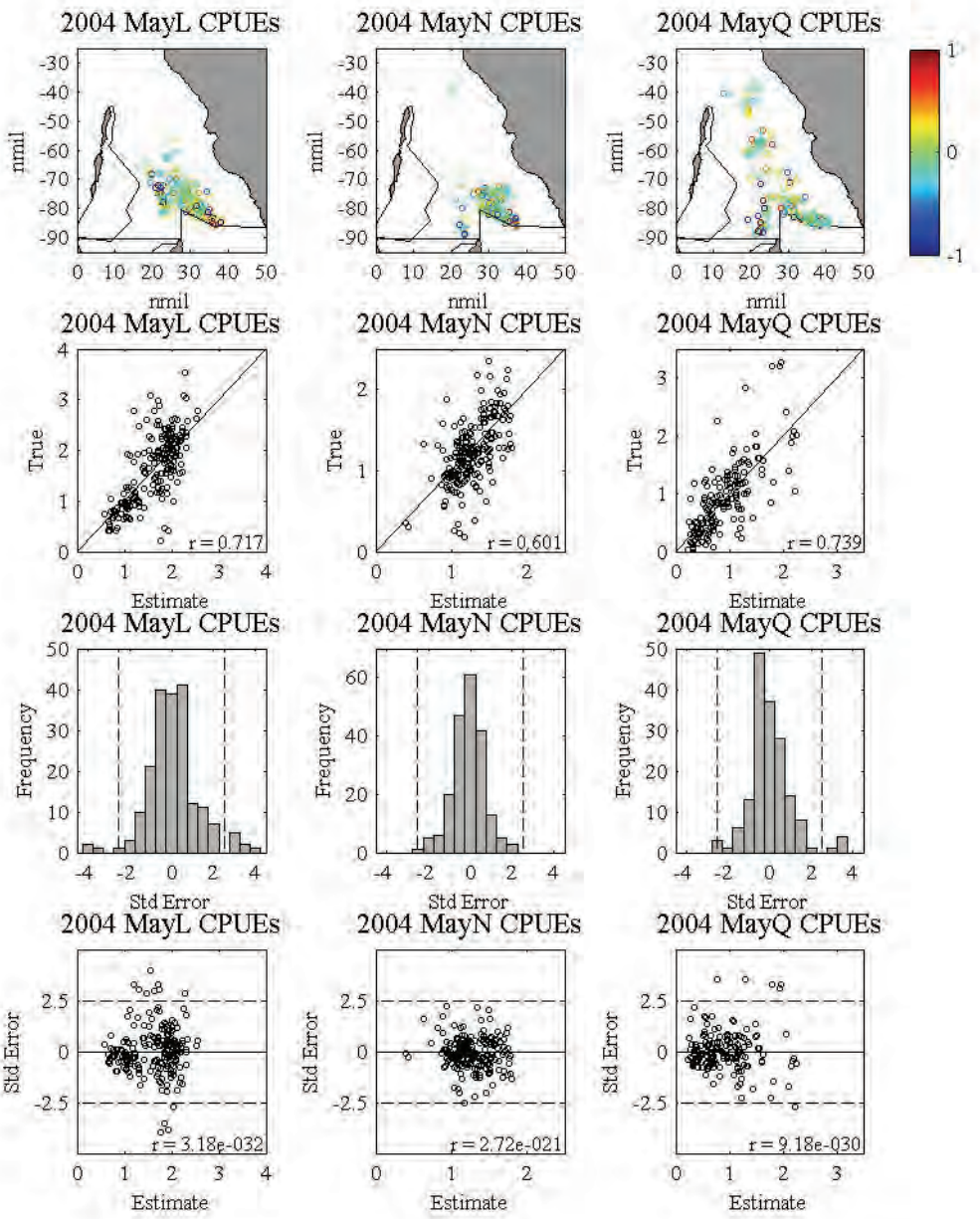


Figure 8.6. *CPUEsvd* values (top row), scatterplot of actual vs estimated values (second row), histogram of standard errors (third row) and standard errors vs estimates (bottom row), May 2004.

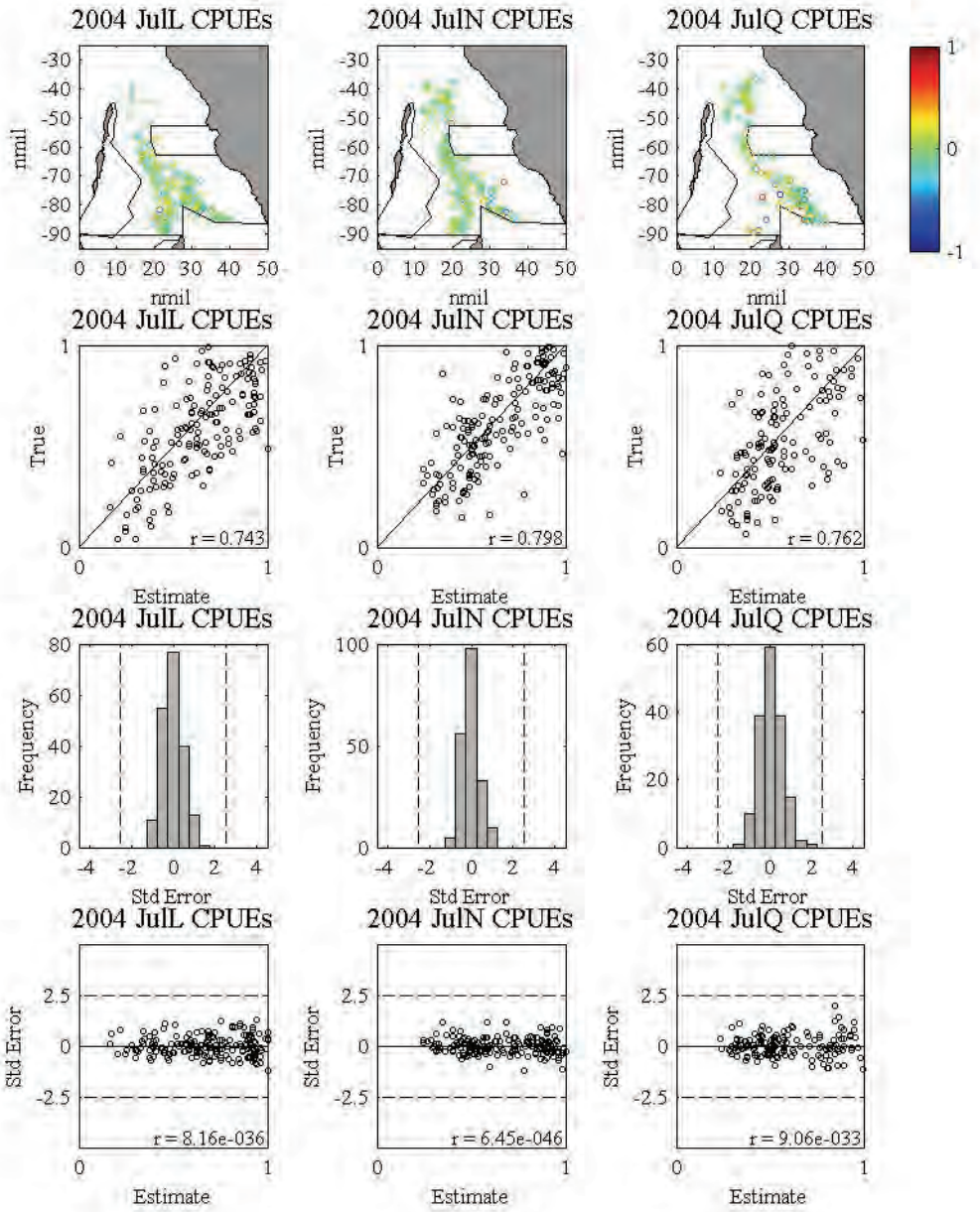


Figure 8.7. *CPUEsvd* values (top row), scatterplot of actual vs estimated values (second row), histogram of standard errors (third row) and standard errors vs estimates (bottom row), July 2004.

The strong positive correlation between the estimates and sample values seen in May and July is evident in many of the other weeks of the 2004 season (Figure 8.3, second row). However, there are also some weeks that show weak positive correlation and one that shows little correlation (AugN). The low correlation seen in AugN is most likely due to the small number of data points in that week.

The mean errors of the 2004 season are quite small in magnitude and centred around the zero line with the largest magnitudes seen in August to October (Figure 8.2, top row). The mean squared errors are also quite small varying between 0 and 0.25 across the season. The errors appear quite randomly distributed from March to August however there is some evidence of a

pattern in the months of September and October which may be attributed to the classical decomposition temporal trend model used to calculate the standardised catch rate values.

The mean squared deviation ratio (Figure 8.8) corresponds to the ratio between the experimental and theoretical variances and shows that the cross validation estimates of the lunar weeks for the 2004 season typically underestimate the theoretical variance, as expected using an ordinary kriging process. There was no pattern across the season for these mean squared deviation ratios. The cross validation results indicate that the spatiotemporal model of Figure 8.1 is adequate for spatiotemporal modelling of the standardised king prawn catch rate in the 2004 season.

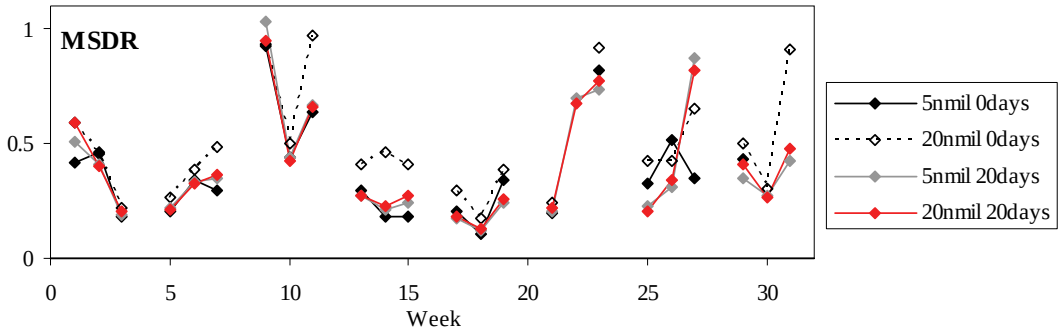


Figure 8.8. Mean squared deviation ratio, cross validation estimates 2004.

8.3.1 KING PRAWN CATCH RATE CROSS VALIDATION

The classical decomposition temporal trend model is combined with the (kriging) estimates of the standardised king prawn catch rate to obtain estimates of the king prawn catch rate. The temporal trend model calculated for the combined 2001-2003 season is given by

$$M_{T,p}^{C123}(t) = m_T^{C123} s_P^{C123} \tag{8.1}$$

where

$$m_T^{C123} = 0.0070T^3 - 0.345T^2 + 3.693T + 28.266\,,\ (T \text{ week number of season}) \tag{8.2}$$

denotes the classical decomposition multiplicative temporal trend model, and

$$s_P^{C123} = \begin{cases} 1.158, & P = L \\ 1.141, & P = N \\ 0.772, & P = Q \end{cases},\ (P \text{ moon phase}) \tag{8.3}$$

denotes the seasonal index.

The cross validation estimates of the *CPUEs* variable for the 2004 season obtained previously can be transformed to *CPUE* values using the trend model of the combined 2001-

2003 season. Post plots of the cross validation estimates of the king prawn catch rates for May and July (Figure 8.9 and Figure 8.10) show that the sample values have been reproduced. The mean and median of each weekly estimation set is similar to the corresponding value of the weekly sample data set (Table 8.4). The variance of each weekly estimation set is considerably smaller than the associated weekly sample data set. The minimum/maximum of each weekly estimation set is higher/lower than the minimum/maximum value of the corresponding weekly sample data set. This behaviour is similar for the remainder of season 2004 (APPENDIX J).

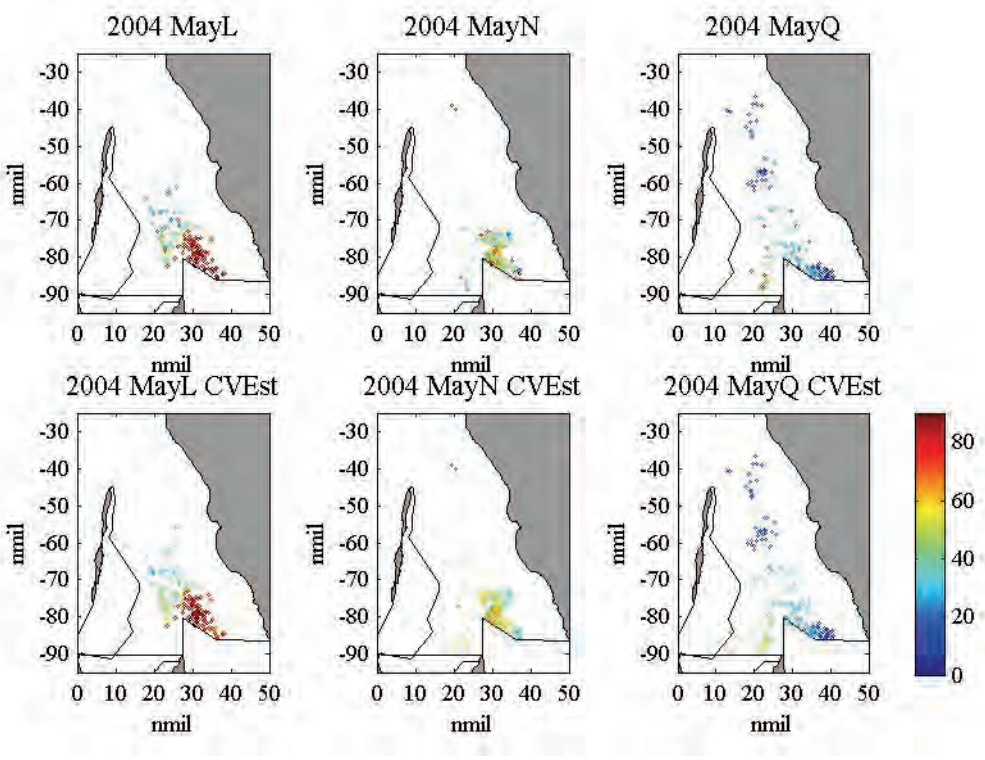


Figure 8.9. *CPUEvd* values (top row) and cross validation estimates (bottom row) for lunar weeks of May 2004.

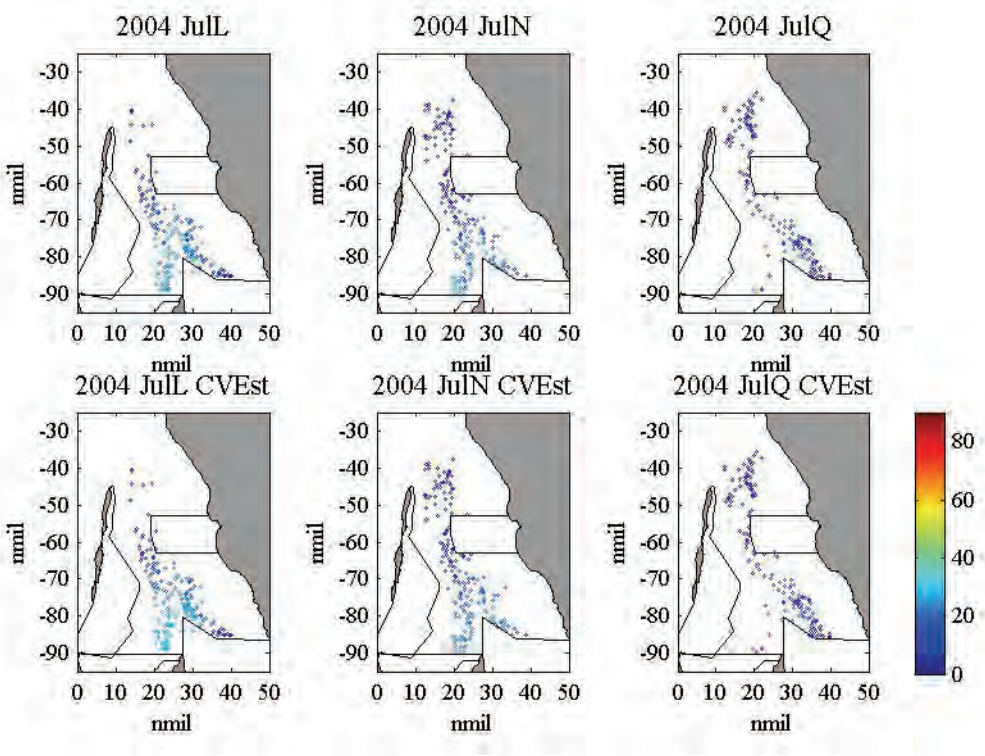


Figure 8.10. *CPUEvd* values (top row) and cross validation estimates (bottom row) for lunar weeks of July 2004.

Table 8.4. Summary statistics of *CPUEvd* cross validation estimates, May & July 2004.

	MayL		MayN		MayQ		JulL		JulN		JulQ	
	Sample	CV Est	Sample	CV Est	Sample	CV Est	Sample	CV Est	Sample	CV Est	Sample	CV Est
Mean	69.57	68.64	46.31	47.69	25.66	25.49	21.16	21.47	17.15	17.33	11.45	11.36
Med	71.62	74.93	45.75	46.46	23.60	22.66	20.63	22.04	16.78	17.66	9.80	9.96
Var	928.13	527.65	216.23	92.85	298.63	171.57	84.96	50.84	51.65	36.20	41.12	20.40
Skew	0.12	-0.17	0.11	-0.09	1.29	0.93	0.04	-0.39	0.32	0.05	0.82	0.56
Min	9.25	25.72	7.11	14.71	1.00	6.21	1.17	5.00	3.68	5.98	1.10	4.12
Max	157.21	113.98	88.86	67.73	91.50	62.74	44.49	33.38	36.64	30.59	28.97	22.95
N	196	196	203	203	165	165	197	197	202	202	166	166

The mean errors of the cross validation estimates were encouraging, ranging between -0.94 and 1.61. However, the mean squared errors ranged from 13.81 to 365.30 with the first half of the 2004 season showing most of the higher values. As the king prawn CPUE estimate is simply a multiple of the standardised estimate, the mean square deviation ratio for the May and July king prawn *CPUE* cross validation estimates (Figure 8.11) are equal to those of the standardised king prawn catch rate cross validation estimates (Figure 8.8).

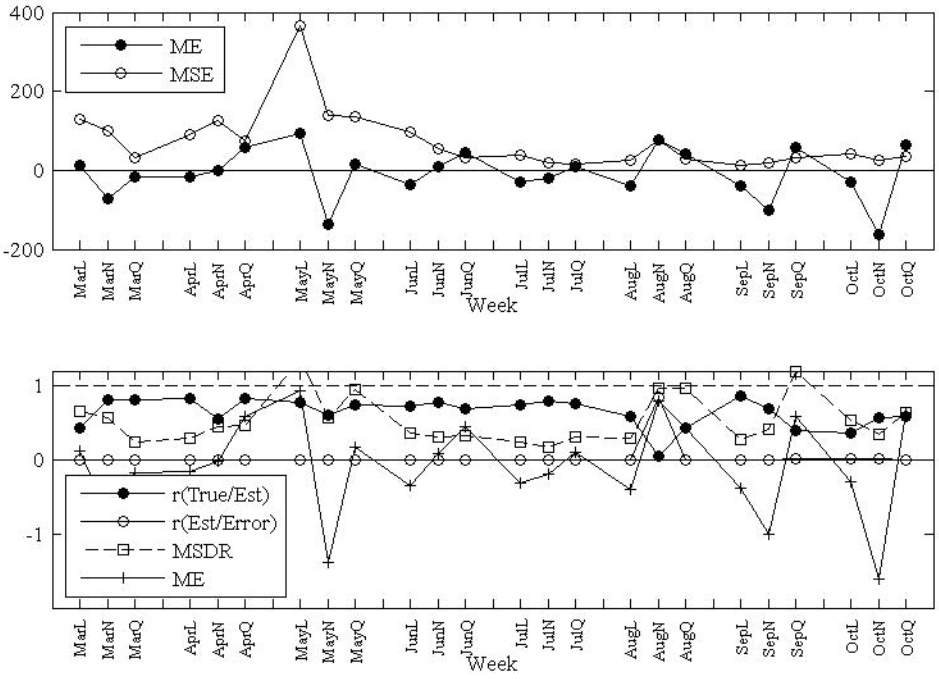


Figure 8.11. Error statistics for *CPUEvd* cross validation estimates for lunar weeks of 2004.

The cross validation process has confirmed that the spatiotemporal model combined with the classical decomposition trend model is adequate for estimation of the king prawn catch rate in the 2004 season. Summary statistics of the sample values are adequately replicated and error statistics are acceptable.

8.4 CATCH RATE ESTIMATION BY FORWARD PREDICTION

In this section we use the spatiotemporal model in an extrapolation mode to predict catch rates for future weeks of the 2004 season. Ordinary spatiotemporal kriging is conducted using the spatiotemporal model of continuity derived from the 2001-2003 data and the search neighbourhood found suitable by the previous cross validation. The input data are the standardised catch rates for the 2004 season up to but not including the lunar month to be estimated. Jacknife estimates were produced for all data locations along with grid estimates calculated on a 1 nmil by 1 nmil grid for all months of the 2004 season except March using the data from the previous month.

Jacknife estimation refers to estimating a set of data values from another non-overlapping data set. This can be implemented to compute estimates for the locations of the actual 2004 data for all months of the 2004 season except March using the data from the previous lunar month. As for the cross-validation estimates, the jacknife estimates can be compared with the actual 2004 data to assess their performance.

The kriging procedure was performed using the standardised king prawn CPUE data for the relevant 2004 month (standardised by the temporal trend model of the combined 2001-2003 season) and the spatiotemporal semivariogram model for the combined 2001-2003 season. The standardised estimates were multiplied by the classical decomposition temporal trend model to obtain jacknife and grid estimates of the king prawn CPUE. As each week involves several days, an average king prawn catch rate estimate was calculated for the 1 nmil by 1 nmil grid locations. Although eight (lunar) months were estimated, results will be shown in this section for the months of May and July, with the results for all other months contained in APPENDIX K.

The temporal search radius suggested by cross validation was 20 days, equal to the temporal range of the spatiotemporal semivariogram, so the input data was limited to include only the standardised catch rate data of the lunar month preceding the lunar month being estimated. The temporal search radius was increased to 30 days for estimation in the second and third week of the lunar month to enable data from the previous lunar month to be used for estimation in that week. Due to the relatively short temporal range of the semivariogram model, the second and third weeks (new moon phase and first quarter moon phase) of each month were re-estimated using the first week of that month (last quarter moon phase) as an additional week of input data.

8.4.1 JACKNIFE AND GRID ESTIMATES

The jacknife estimates underestimate the high catch rates evident in the first two weeks of May and fail to estimate at all sample locations (Figure 8.12). The mean, median and variance of the estimated values are lower than those of the sample values (Table 8.5). The minimum/maximum of the estimated values is much higher/lower than the corresponding sample values. The number of locations at which estimates are made decreases as the distance in time between the available data and prediction location increases. A larger search radius could possibly address this issue but the additional data would be quite far removed and would potentially result in over smoothing.

The mean grid estimates fail to capture the behaviour of the actual catch rates in the ENA during the weeks of May (Figure 8.12) due to the lack of available data in that region in April. The grid estimates show a much lower mean, median and variance than the corresponding actual catch rates.

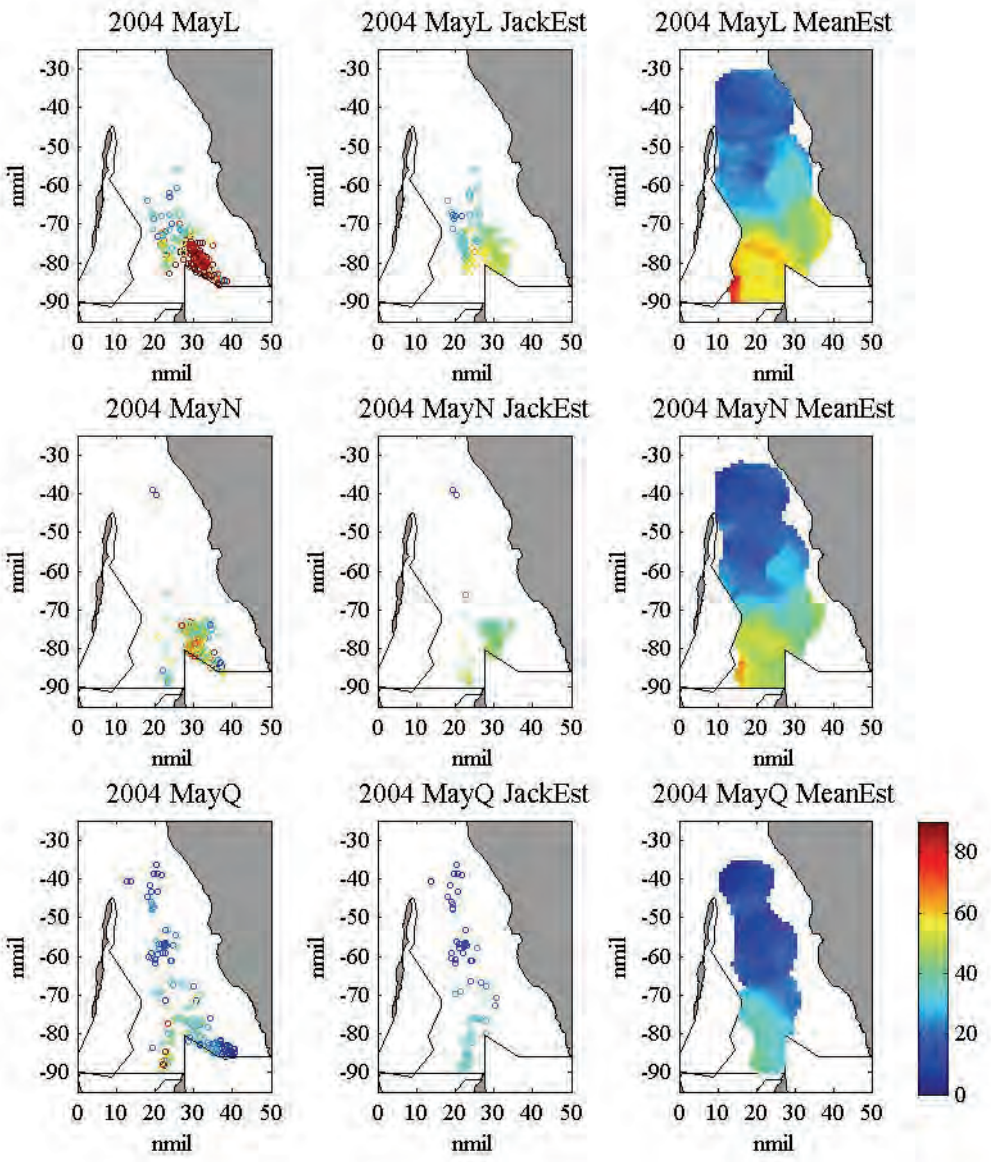


Figure 8.12. *CPUEvd* data (left), jacknife estimates (centre) and grid estimates (right) for last quarter (top row), new moon (middle row) and first quarter (bottom) lunar weeks of May 2004 using data from April 2004.

Table 8.5. Summary statistics of *CPUEvd* jacknife estimates for lunar weeks of May 2004.

	MayL		MayN			MayQ		
	Sample	Jack Est	Sample	Jack Est	Jack Est	Sample	Jack Est	Jack Est
Mean	69.573	45.992	46.312	42.422	68.571	25.659	22.653	38.962
Med	71.616	47.785	45.752	43.027	70.592	23.600	20.868	44.141
Var	928.133	102.907	216.229	28.217	75.156	298.634	116.795	215.026
Skew	0.124	-0.786	0.112	-2.749	-2.966	1.285	0.299	-0.946
Min	9.249	17.079	7.108	12.821	12.821	0.999	5.399	5.399
Max	157.211	61.782	88.864	52.416	80.095	91.501	41.064	58.519
N	196	161	203	135	203	165	63	159

Table 8.6. Summary statistics of *CPUEvd* grid estimates for lunar weeks of May 2004.

	MayL		MayN			MayQ		
	Sample	Grid Est	Sample	Grid Est	Grid Est (inc MayL)	Sample	Grid Est	Grid Est (inc MayL)
Mean	69.573	34.479	46.312	28.923	38.505	25.659	19.657	33.805
Med	71.616	29.286	45.752	24.36	32.902	23.600	17.021	37.959
Var	928.133	240.219	216.229	185.075	412.962	298.634	97.211	214.504
Skew	0.124	0.489	0.112	0.477	0.172	1.285	0.498	-0.365
Min	9.249	8.648	7.108	6.901	6.901	0.999	4.979	4.979
Max	157.211	80.103	88.864	71.007	79.379	91.501	45.232	60.924
N	196	9392	203	7437	9018	165	2330	5522

The MayN and MayQ jacknife estimates incorporating the additional MayL input data produced estimates at more sample locations than the previous estimates but cannot produce the sample statistics (Figure 8.13, Table 8.5). The mean and median of the revised estimates are much higher than the previous estimates and the sample values. The variances of the revised estimates are higher than the previous estimates but still lower than the sample values. The minimum values of the revised estimates are equal to those of the original estimates but the maximum values are considerably higher. The revised mean grid estimates for MayN and MayQ involving the additional week of input data overestimate the values of the actual catch rates in the ENA. The revised estimates show much a higher mean, median and variance than the previous grid estimates with similar minima and maxima.

The jacknife estimates of July appear to quite accurately represent the actual catch rates evident in the three weeks of July (Figure 8.14). All sample locations are estimated in JulL and all but four locations are estimated in JulN. However in JulQ only 55 of 166 sample locations are estimated by the jacknife process. The mean, median, variance and minimum of the estimated values are slightly higher than those of the sample values (Table 8.7) while the maximum is slightly lower.

The mean grid estimates for the three weeks of July appear to reproduce the areas of high and low catch rates evident in the actual July catch rates with the exception of an area of high values above the TPSA not visible in the sample values (Figure 8.15). The mean, median, minimum and maximum of the July grid estimates are indicative of the corresponding sample data (Table 8.8).

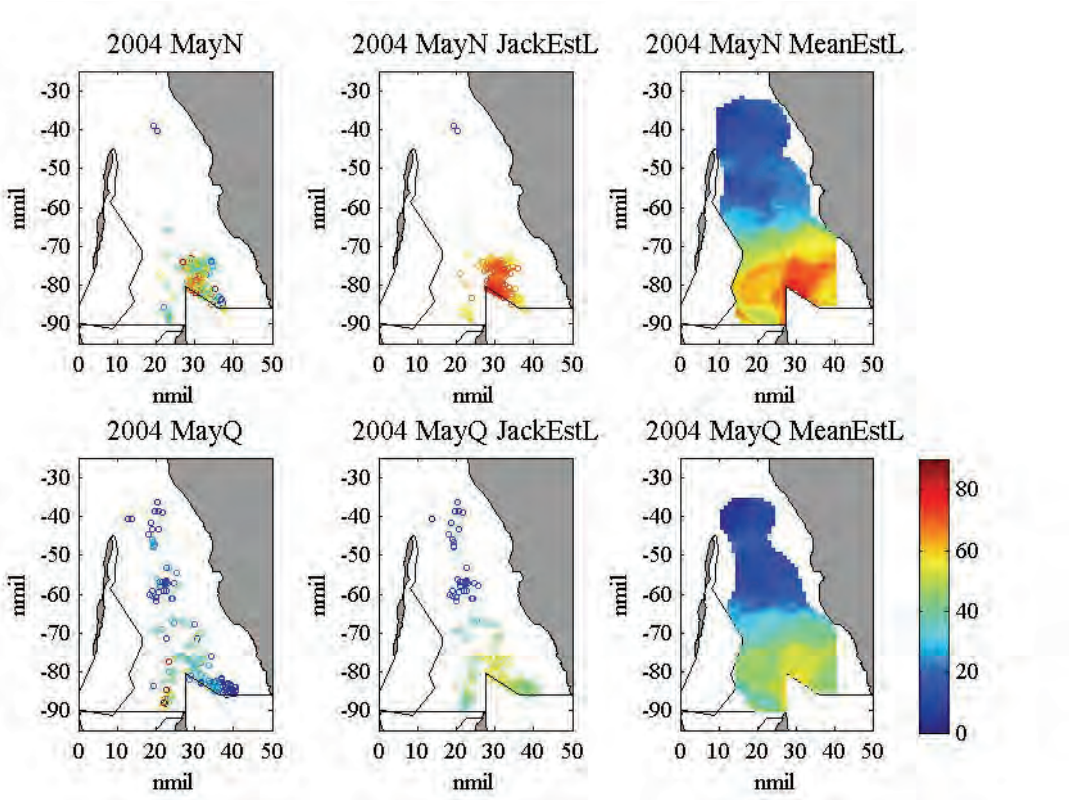


Figure 8.13. *CPUEvd* data (left), jacknife estimates (centre) and grid estimates (right) for new moon (top row) and first quarter (bottom) lunar weeks of May 2004 using data from April and last quarter lunar week of May 2004.

Table 8.7. Summary statistics of *CPUEvd* jacknife estimates, July 2004.

	JulL		JulN			JulQ		
	Sample	Jack Est	Sample	Jack Est	Jack Est (inc JulL)	Sample	Jack Est	Jack Est (inc JulL)
Mean	21.163	24.128	17.148	19.048	15.918	11.451	11.771	9.392
Med	20.626	25.696	16.775	19.264	14.999	9.797	13.012	9.025
Var	84.964	56.023	51.648	28.044	26.515	41.121	14.768	9.493
Skew	0.043	-1.052	0.316	-0.015	0.287	0.818	-0.803	0.456
Min	1.165	4.145	3.677	4.142	5.833	1.096	1.147	3.822
Max	44.485	36.456	36.641	31.273	25.072	28.971	18.537	17.912
N	197	197	202	198	202	166	55	164

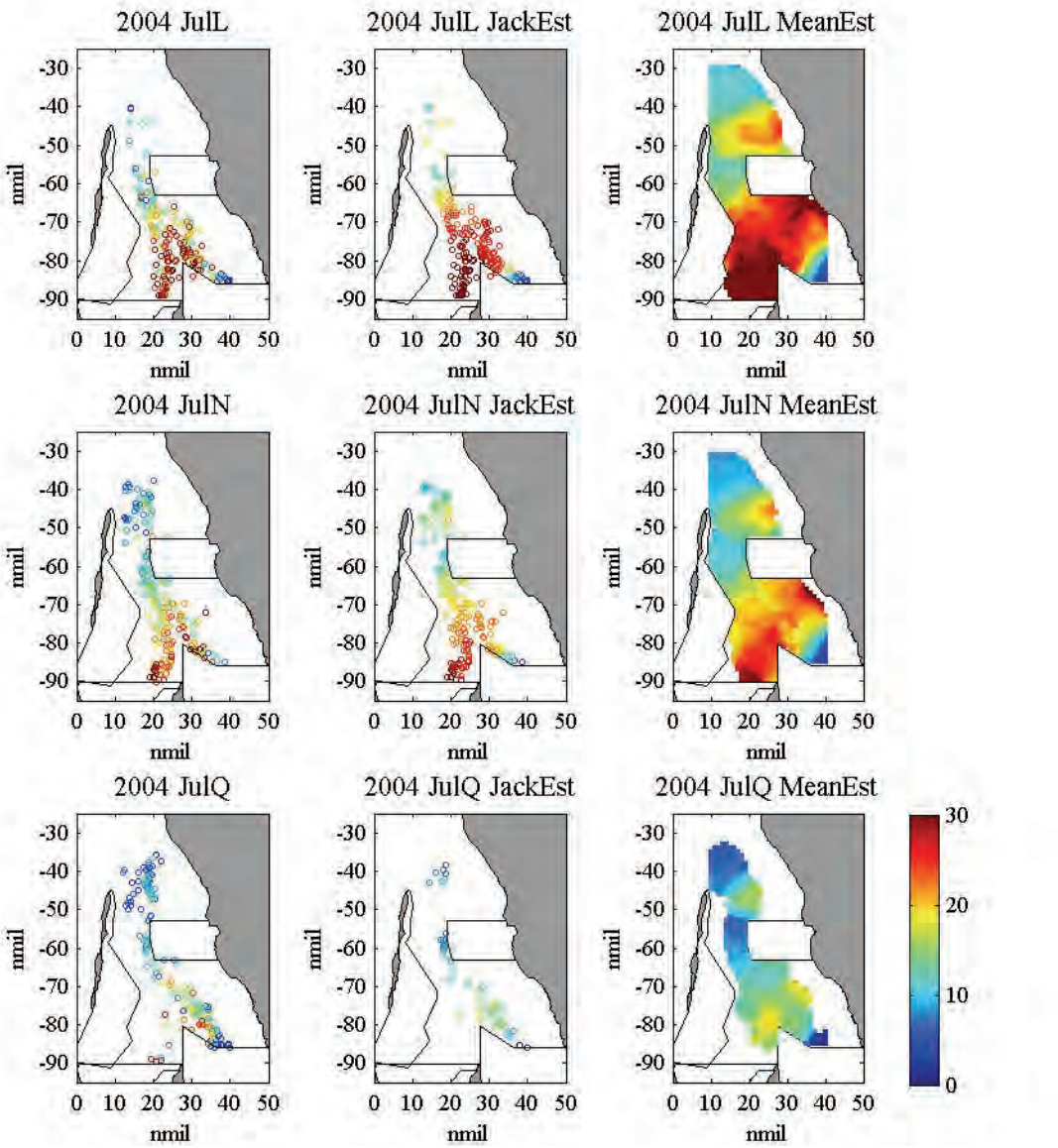


Figure 8.14. *CPUEvd* data (left), jacknife estimates (centre) and grid estimates (right) for last quarter (top row), new moon (middle row) and first quarter (bottom) lunar weeks of July 2004 using data from July 2004.

Table 8.8. Summary statistics of *CPUEvd* grid estimates, July 2004.

	JulL		JulN			JulQ		
	Sample	Grid Est	Sample	Grid Est	Grid Est (inc JulL)	Sample	Grid Est	Grid Est (inc JulL)
Mean	21.163	21.257	17.148	17.082	14.829	11.451	11.841	10.581
Med	20.626	21.185	16.775	17.818	14.386	9.797	12.439	10.632
Var	84.964	55.679	51.648	31.959	28.711	41.121	14.437	14.516
Skew	0.043	0.109	0.316	-0.039	0.367	0.818	-0.093	0.339
Min	1.165	3.383	3.677	2.353	5.416	1.096	1.164	3.805
Max	44.485	43.893	36.641	35.292	25.685	28.971	23.071	18.276
N	197	8450	202	6936	8180	166	1820	6370

The JulN and JulQ estimates incorporating the additional JulL input data produced estimates at more sample locations than the previous estimates but offer little improvement to more accurately produce the sample statistics (Figure 8.15, Table 8.7). The revised estimates

have slightly lower means, medians, variances and maxima than the corresponding previous estimates as well as slightly higher minima.

The use of the JulL data as input data appears to improve the revised mean grid estimates for JulN and JulQ by removing the area of high catch rates above the TPSA and just below the TPSA seen in the original mean grid estimates. The summary statistics of the revised grid estimates are quite similar to those of the original grid estimates (Table 8.8).

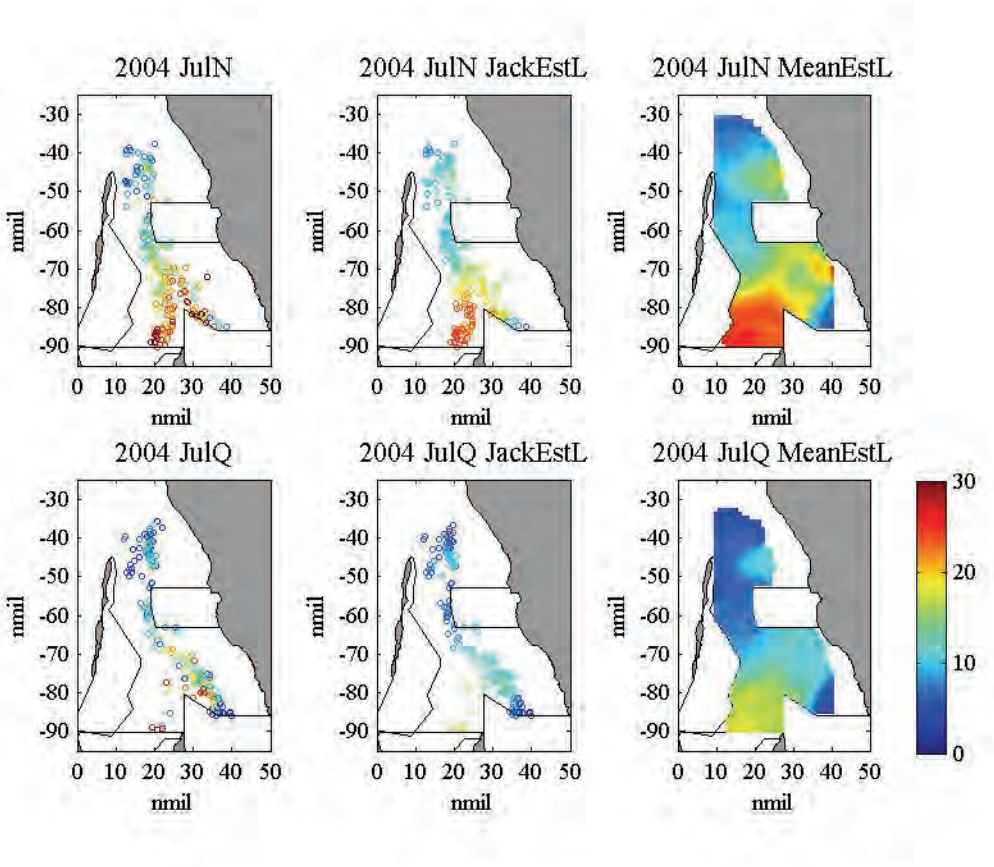


Figure 8.15. *CPUEvd* data (left), jacknife estimates (centre) and grid estimates (right) for new moon (top row) and first quarter (bottom) lunar weeks of July 2004 using data from June and last quarter lunar week of July 2004.

Accuracy measures for the jacknife estimates (Figure 8.16) show that the highest magnitude of mean errors and mean squared errors occur in the month of May, which is the month involving the highest actual catch rates. The mean errors in May are negative for the original estimates showing that the actual catch rates are underestimated. The revised estimates result in positive mean errors for May showing that the actual catch rates are overestimated. The mean standardised errors are also large in magnitude for the month of May, although the months of August through October also show large standardised errors. As for the mean errors, the mean standardised errors are negative for the original jacknife estimates and positive for the revised estimates. The MayL and MayQ weeks of the original estimates along with the MayN

and MayQ weeks of the revised estimates are the only weeks that show mean standard deviation ratios greater than one.

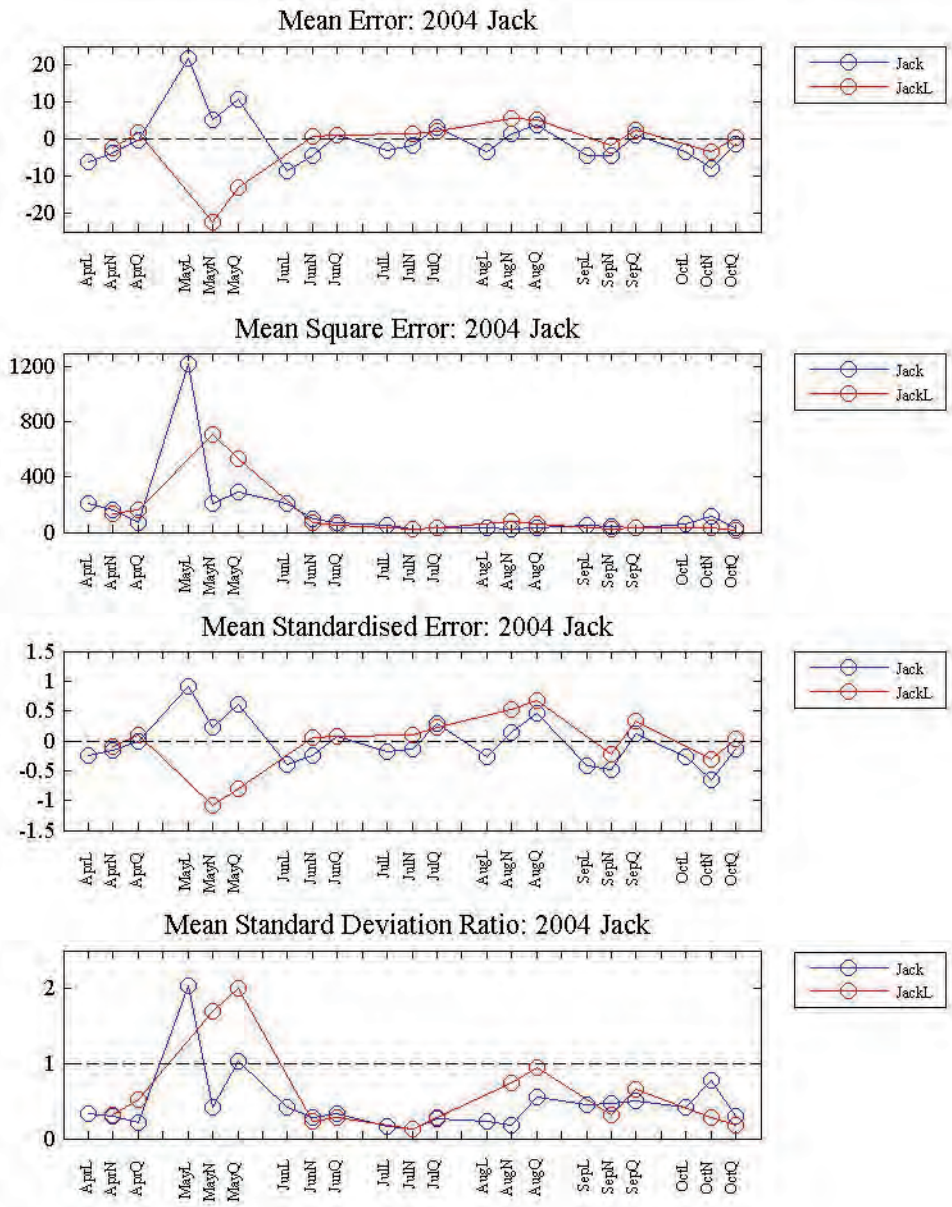


Figure 8.16. Mean error (top row), mean square error (second row), mean standard error (third row) and mean standard deviation ratio (bottom row) for *CPUEvd* jacknife estimates of lunar weeks of 2004.

Correlation values are shown in Figure 8.17 with significant correlations indicated by filled markers. The jacknife estimates show positive correlation with the sample values for most of the weeks estimated. Only AugQ, SepQ, OctL and OctN show little evidence of correlation. The revised estimates do not necessarily show better correlation with the actual catch rate values. The jacknife estimation errors show evidence of correlation with the estimates for MayQ and OctN of the original estimates and for MayQ and OctQ of the revised jacknife estimates.

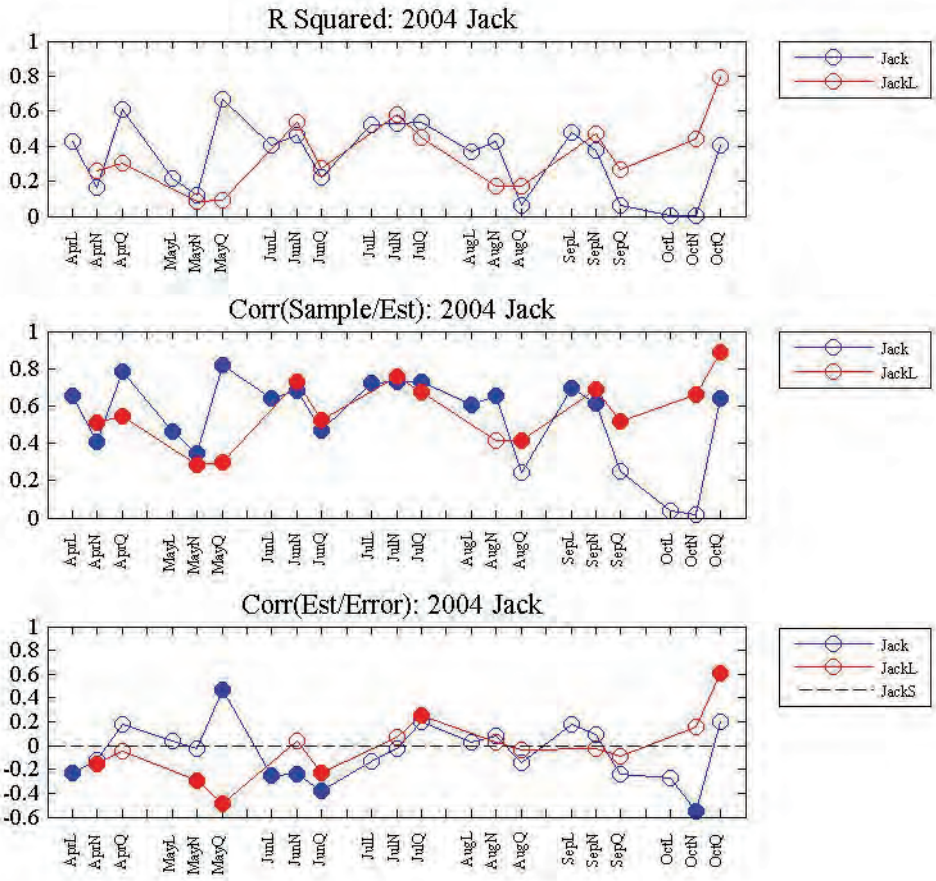


Figure 8.17. Coefficient of determination (top row), correlation of true and estimated values (second row) and correlation of estimated values and errors (bottom row) for *CPUEvd* jacknife estimates of lunar weeks of 2004.

8.5 ESTIMATION INCORPORATING SURVEY DATA

We have shown that it is possible to predict the king prawn catch rate for all months of the 2004 season using a spatiotemporal geostatistical model obtained from the data of seasons 2001 to 2003, along with the logbook catch rate data of the previous months of 2004. However, this method does not adequately capture the relatively high catch rates in the first week of May as the ENA opens to trawling (Figure 8.12). The catch rates in the ENA are relatively high compared to those further away from the ENA and they are significantly higher than the estimate of the classical decomposition model. One possible solution to this problem is to utilise survey data available in the ENA region in the months before it is opened to trawling just before the last quarter moon phase (Denham & Mueller, 2010).

Recall that the survey data consist of 17 locations across the study region (Figure 3.5), which were sampled around the last quarter moon phase in the months of March and April of each season. The March and April survey data for seasons 2001 to 2003 show similarities with the actual catch rates within the ENA in the first week it is open. It was decided that the average

of the two months survey data was the most reasonable indicator of the catch rate values in the ENA (Figure 8.18). And the use of the average survey data in the estimation process would help to reproduce the high catch rate behaviour in the ENA.

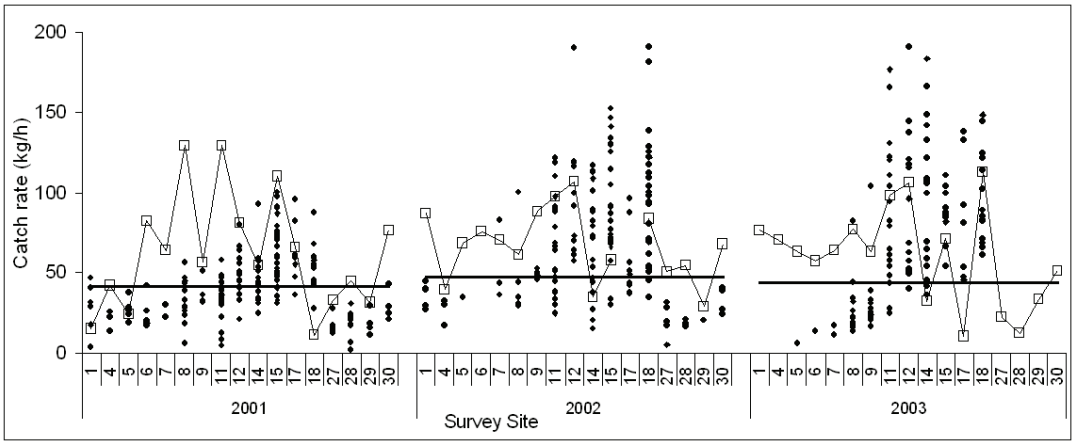


Figure 8.18. Average survey catch rates (white squares), logbook catch rates at survey locations (black circles) and classical decomposition model fit (solid line) for week of last quarter moon phase in season 2001-2003.

As their values were considered indicative of the first week of fishing in May when the ENA opens, the survey data of season 2004 were detrended and deseasonalised using the trend and seasonal index for the last quarter moon phase of May. They were allocated a date from the preceding week to enable their use in the estimation process which was directly affected by the short temporal range of the semivariogram model. Estimates over the fishing region and jacknife estimates for the actual logbook catch rate data locations (Figure 8.19) demonstrate the ability to better capture the high catch rates in the ENA in MayL week.

While inclusion of the survey data improved the estimation for the last quarter, this was not the case for the weeks of the new moon and first quarter moon phase of May 2004. The relatively high catch rates are much fewer in these weeks and the estimates involving the survey data are much higher than those evident in the actual catch rates. The estimates involving no survey data are more representative of the actual catch rates (Figure 8.12).

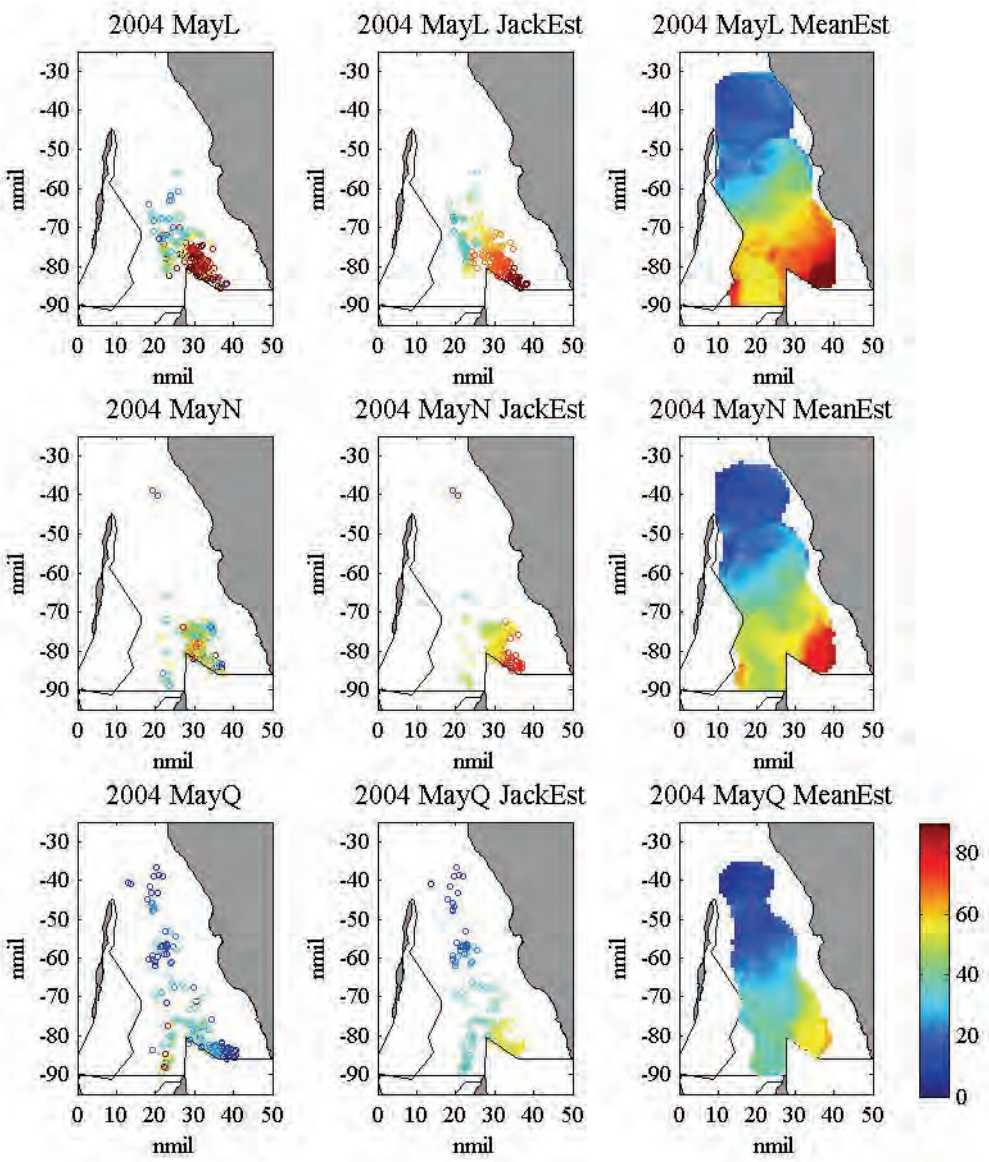


Figure 8.19. CPUE values (left), jacknife estimates (centre) and grid estimates (right) for lunar weeks of May 2004. Estimates made using survey data as additional input data.

Accuracy measures for the jacknife estimates support the use of the survey data to estimate for the last quarter moon week (Figure 8.21). Estimation using the survey data decreases the magnitude of the errors for the week of the last quarter moon phase but increases the magnitude of the errors for the weeks of the new moon and first quarter moon phase. The R^2 value of the estimates using survey data increases for the last quarter moon phase week but decreases for the weeks of the new moon and first quarter moon phase (Figure 8.20).

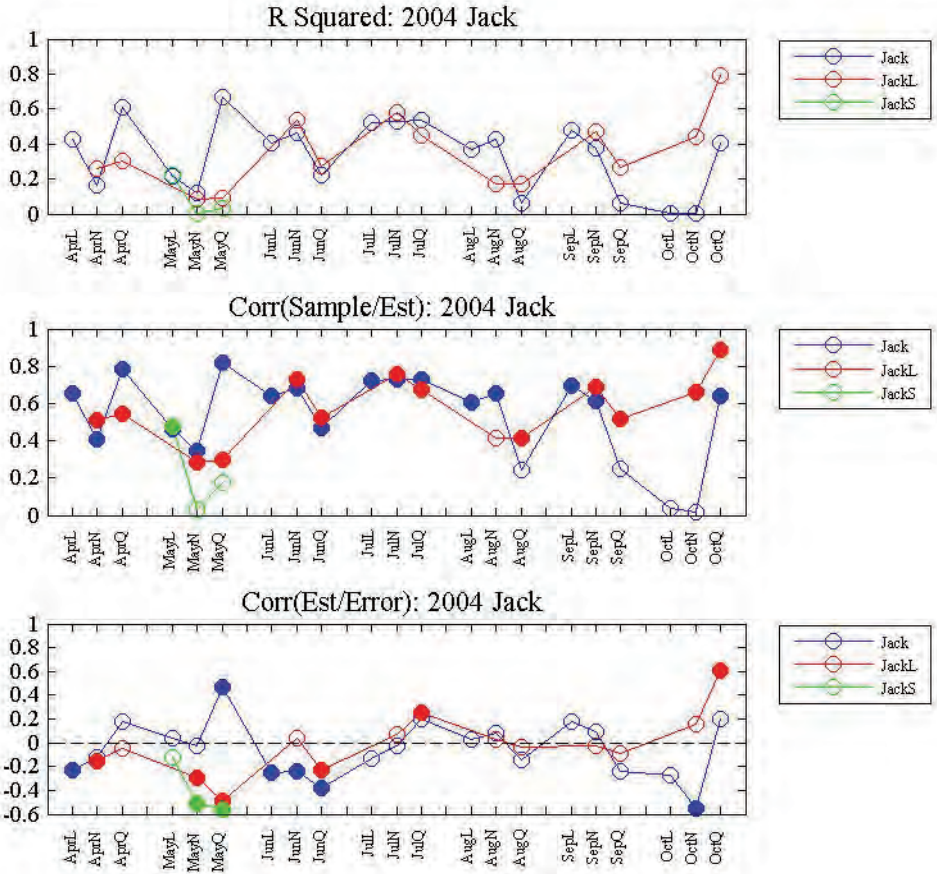


Figure 8.20. Coefficient of determination (top row), correlation of true and estimated values (second row) and correlation of estimated values and errors (bottom row) for *CPUEvd* jacknife estimates of lunar weeks of 2004 using data from previous month (blue) along with additional data from last quarter lunar week of estimation month (red) or additional survey data (green).

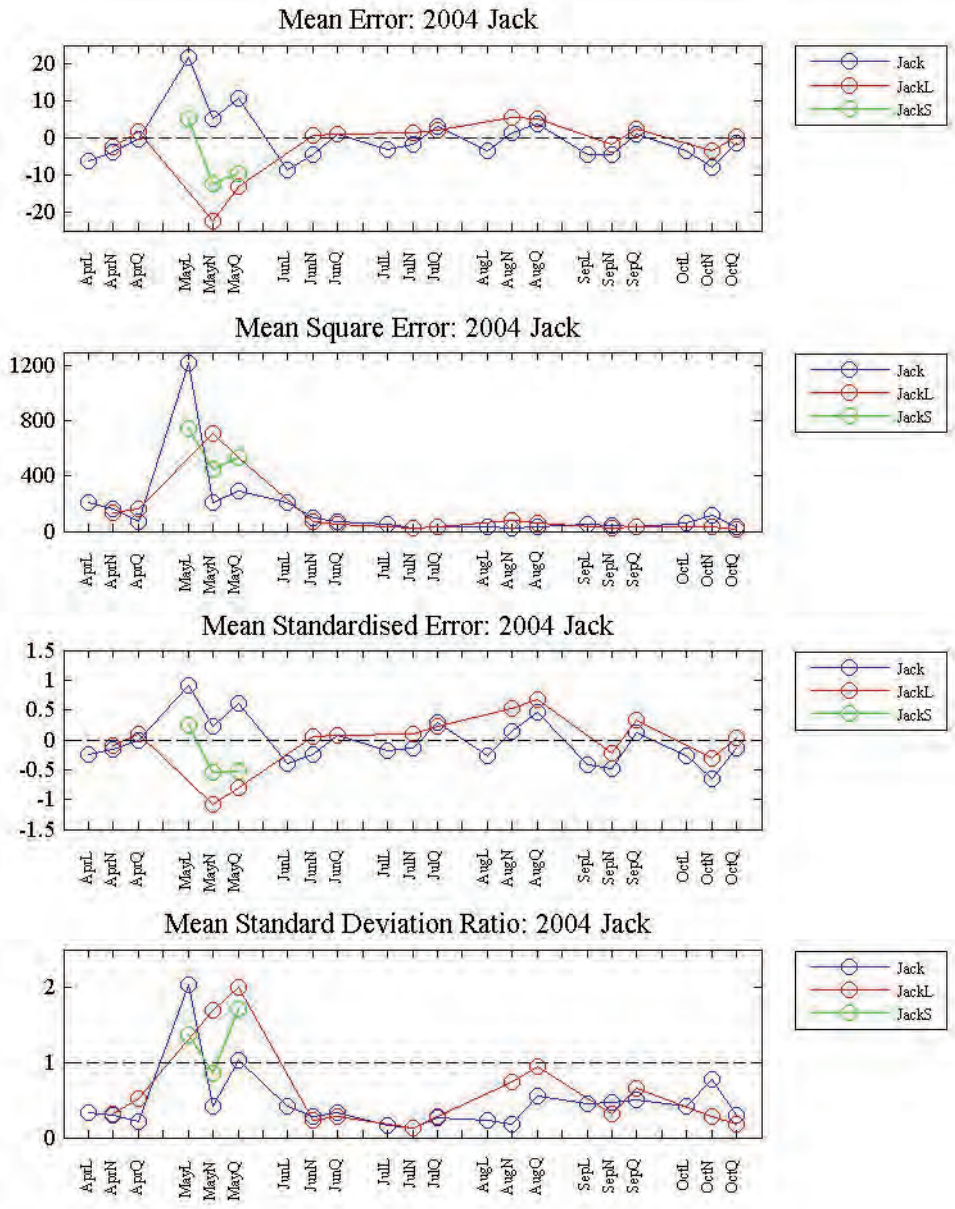


Figure 8.21. Mean error (top row), mean square error (second row), mean standard error (third row) and mean standard deviation ratio (bottom row) for *CPUEvd* jacknife estimates of lunar weeks of 2004 using data from previous month (blue) along with additional data from last quarter lunar week of estimation month (red) or additional survey data (green).

8.6 CHAPTER SUMMARY

Cross validation verified the use of the spatiotemporal semivariogram model of the *CPUEsvd* data to predict catch rates at unsampled locations via a weighted estimation (kriging) procedure. The inclusion of temporal information was beneficial to the performance of kriging estimates for season 2004, particularly in the absence of closely related spatial information. Building on this, estimates of 2004 catch rates were determined using data from the previous month as conditioning data. This enabled estimates to be made successively up to two weeks in advance, unless the region involved the opening of a fishing region resulting in significantly different catch rates than seen in the conditioning data. Estimates for the third week of a lunar month

involved temporal separations too large to use the previous month's data as meaningful conditioning. By including the first week's data, estimates could be improved for this third week unless the first week of the month also involved the opening of a fishing region.

Survey data obtained as part of the annual recruitment surveys were used as supplementary information when the region/period being estimated involved the opening of the ENA to fishing. The fishery-independent survey data served as more appropriate conditioning data for that region. However, their impact was seen to persist in the following weeks when in reality the impact of the opening subsided over time. Therefore, it may be appropriate to scale the influence of the survey data by a multiplicative factor in relation to the number of days from opening.

CHAPTER 9

SPATIOTEMPORAL SIMULATION

9.1 INTRODUCTION

In the previous chapter, estimation of the king prawn catch rate, using data up to the previous lunar month, provided smoothed estimates over the region. There was good correlation in many lunar weeks between the estimates made for the 2004 season using the data, and associated models, from the previous three seasons with the actual catch data for the lunar weeks in 2004. To further develop the model of the catch rates, simulation was used to investigate and model the local variability of the catch rate estimates. Geostatistical simulation offers an advantage over kriging. As kriging is based on a local average of the data, it produces output more smoothed than the real phenomenon and only one value for every spatiotemporal location in the study area. The ability to make informed decisions on the catch rates over space and time is limited by this single smoothed estimation surface. Simulation is an alternative approach that produces potentially more realistic representations of the local variability than the (kriging) estimation process by adding local variability back into the realisations it generates that is lost in kriging. The variability that simulation realisations add to the predicted value at a particular location has a mean of zero, so that the average of many simulated realisations tends toward the kriging prediction.

A geostatistical simulation process was used in this study to further refine areas of known high or low catch rates, along with other areas involving a higher degree of variability. As the simulation process to be used involves a Gaussian algorithm, normal scores of the standardised king prawn catch rate were used which can be back transformed to obtain simulated standardised catch rate values. This required the use of the spatiotemporal semivariogram of the normal scores (Section 7.4) which displayed similar ranges of spatial and temporal continuity to the semivariogram of the standardised catch rate used for estimation, allowing forward prediction of catch rates into the next lunar fishing month. As the collection of data moves to electronic means in Australian fisheries, leading to faster data entry and verification, the ability to forecast the level of uncertainty of catch rates for the next month of fishing gives scientists a useful tool for the management of their fishery.

9.2 SIMULATION PROCESS

Stochastic simulation is the process of building alternative, equally probable, high-resolution models of a variable's distribution. It is a different approach to the problem of estimating the value of a variable at a location than kriging. Kriging derives an estimate and then attaches to it a confidence interval (via the kriging variance). The simulation process first assesses the uncertainty about the variable at a location dependent on the information available. It then derives a possible realisation for that location based on the model of uncertainty at that location (*local* uncertainty). The problem of generating joint realisations at all locations in a spatiotemporal domain is made simpler by the sequential simulation process. A realisation is made at each location conditional to not only to the original sample data but also the previously simulated data. Multiple realisations can be produced for the same spatiotemporal domain, providing a visual and quantitative measure of the spatiotemporal uncertainty.

Implementation of the sequential principle under the multiGaussian RF model is referred to as sequential Gaussian simulation (SGS). A random path is defined that visits each node of the grid to be estimated exactly once. At each node, a SK procedure is used to estimate the mean and variance of the Gaussian ccdf of the normal scores, using the spatiotemporal variogram model of the normal scores and a given search neighbourhood. Under the multiGaussian model, the mean and variance of the ccdf at a location are identical to the SK estimate and SK variance obtained from the conditioning data. The conditioning information consists of local normal score data and values simulated at previously visited nodes. A simulated value is calculated by drawing a random value from a normal distribution with a mean of zero and a variance equal to the local kriging variance. This value is added to the kriged estimate to produce a simulated value. This simulated value is added to the data set and the process repeated at the next node. Once all nodes are estimated, the simulated normal values are then back-transformed to the original variable.

9.3 SIMULATION OF 2004 KING PRAWN CATCH RATE

Simulation of the king prawn *CPUE* values was conducted for all months of the 2004 season except March using the data up to and including the previous month and the normal scores semivariogram model obtained from the data of seasons 2001 to 2003. In the previous chapter, estimation was performed on the *CPUEs* data, which was then combined with the temporal trend model to obtain estimates of the king prawn catch rate. In a similar manner, simulated king prawn catch rate values were obtained by first calculating simulated values of the

normalised standardised catch rate values (*CPUEsN*), which were back-transformed to standardised catch rates and then combined with the temporal trend model to produce simulated catch rate values. The classical decomposition temporal trend model used to transform the adjusted simulated values was outlined previously (8.2 - 8.5).

Eight months (March-October) involving three weeks (L, N, Q) in each were simulated in 2004. As for the estimation process, simulations are not made for the weeks corresponding to the full moon as the fishing region is typically closed during these periods. For each lunar month, simulated catch rate values were calculated for each day of the lunar month on a 1 nmil by 1 nmil grid within the fishing region. Simulation at each grid node uses a combination of the closest data values along with previously simulated values.

The spatiotemporal semivariogram model used in the simulation procedure is that of the normal scores of the standardised catch rate for the combined 2001-2003 season (Table 9.1, Figure 9.1). A modified GSLIB program *sgsimtemp.exe* was developed and used along with the associated parameter file (APPENDIX G.4). The search neighbourhood used in the simulation procedure was inferred from the previous estimation procedure. This involved a search window with a radius of 20 nmil and 30 days, along with a minimum of 4 sample data and a maximum of 20 required to enable estimation. The temporal search radius meant that only data from the previous month were used in the actual simulation process. The number of simulated nodes to use in any simulation was set to 6. This number was set relatively low to prevent over-smoothing of the simulated values or screening of more relevant sample data.

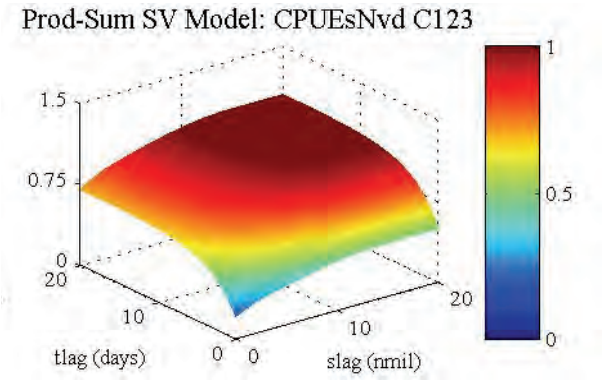


Figure 9.1. Product-sum semivariogram models of *CPUEsNvd* for combined season 2001-2003.

Table 9.1. Product-sum semivariogram model parameters of *CPUEsNvd* for combined season 2001-2003.

Marginal Spatial Model			Marginal Temporal model		
Nugget		0.20	Nugget		0.00
1st structure	Type	spherical	1st structure	Type	exponential
	Range	3		Range	10
	Sill	0.04		Sill	0.52
2nd structure	Type	spherical			
	Range	15		Global Sill	1.00
	Sill	0.3			

Back-transformation of the normal score simulated values to standardised catch rate values was performed using linear interpolation to the lower limit zero, linear interpolation for the middle classes and hyperbolic model extrapolation with parameter $\varpi = 1.5$ to the upper limit which varied according to the lunar week of simulation. This upper limit was set to the average maximum value of the standardised catch rate for the 2001 to 2003 seasons for the corresponding week (Table 9.2). For JunL, only the average of seasons 2001 and 2002 was used, as season 2003 had an exceptionally high maximum for this lunar week.

Table 9.2. Hyperbolic parameter for upper limit of normal score back-transformation.

Phase/Month	April	May	June	July	August	September	October
L	2.5	3.6	2.1	1.6	1.8	2.4	2.5
N	2.2	3.2	1.8	1.7	2.4	2.9	1.7
Q	2.5	2.6	2.0	2.2	2.8	2.8	2.0

9.3.1 SIMULATION OF JULY 2004 CATCH RATE VALUES

As estimates obtained via kriging in the previous chapter for the lunar month of July 2004 were shown to quite accurately reproduce the behaviour of the actual catch rates recorded in that month (Figure 8.13 and Figure 8.14), the results of the simulation process for the July lunar month are presented first rather than those for the month of May. The mean grid estimates for the three weeks of July reproduced the areas of high and low catch rates evident in the actual July catch rates with the exception of an area of high catch rates above the TPSA not present in the sample values. Simulated catch rate values were investigated to model the variability within the region for the July lunar month of fishing. Simulated values for a particular day, rather than an average over a lunar week, were the initial focus to investigate the variability presented by simulation. In particular, the results for the first day of each lunar week of Jul are presented.

Comparison of the first simulation of the first day from the three weeks of July 2004 (Figure 9.2, second row) shows the decrease in the amount of locations being simulated at in the successive weeks. They also indicate a general decrease in catch rate values from the first to the

third week in July, as seen in the sample values (first row). The first grid simulation (of 100 simulations) of the first day of each lunar week of July shows much more variability than the corresponding estimated grid values (fourth row) obtained previously via kriging.

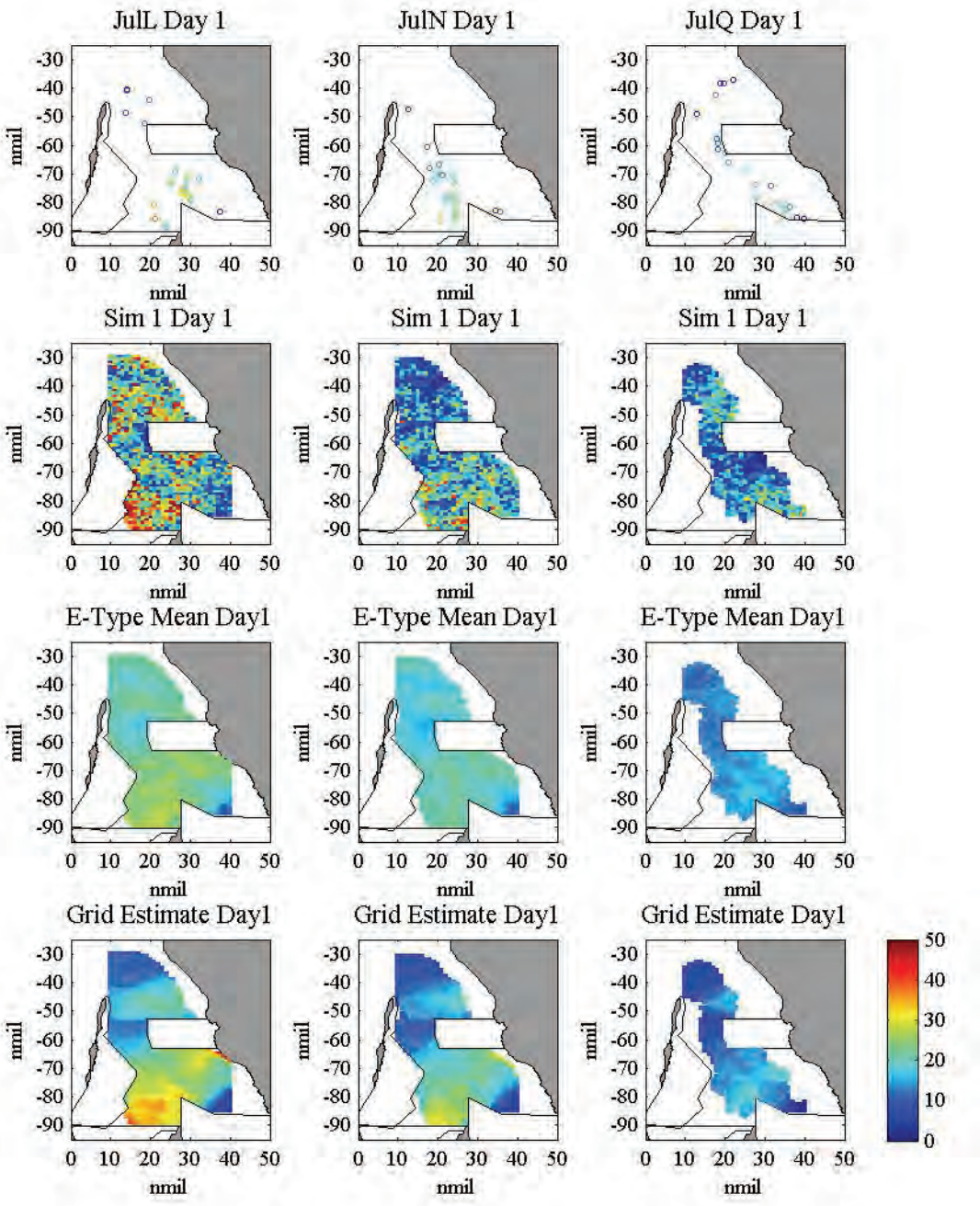


Figure 9.2. *CPUE* values (top row), individual simulation of *CPUE* values (second row), mean of 100 simulations of *CPUE* values (third row) and OK grid estimates of *CPUE* values for the first day of the three lunar weeks of July 2004: JulL (left), JulN (centre) and JulQ (right).

As with kriging, a location estimation map can be obtained using simulation by adding every simulation together and dividing by the number of simulations. For sequential gaussian simulations, this average should tend towards the kriging mean over a large number of simulations. The E-type mean simulated values of the 100 simulations for the first day of JulL

(Figure 9.2, third row) display similar areas of high and low values as the kriging estimates. As for the estimates these mean simulations indicate that the catch rate values are higher in the southern region and lower in the northern region. However, the E-type means show even smoother values than the OK estimates. This means the area of high values seen to the right of the ENA in the estimates is also seen in the simulated values but the values are slightly lower. Similarly the area of low values visible in the northern region in the estimates corresponds to an area of slightly higher simulated values.

Multiple realisations of catch rates over the fishing region were obtained by simulation for a particular day. The first eight realisations of simulated catch rate values for Day 1 of JulL 2004 (Figure 9.3) produce realisations throughout the region covered by the sample data, although there are many locations that overestimate the catch rates evident in the catch data. Most of the realisations show various differing small regions of high values spread across the region. Such features that only occur on a few maps can be regarded as uncertain, but consistent with the model of spatiotemporal continuity used in the simulation.

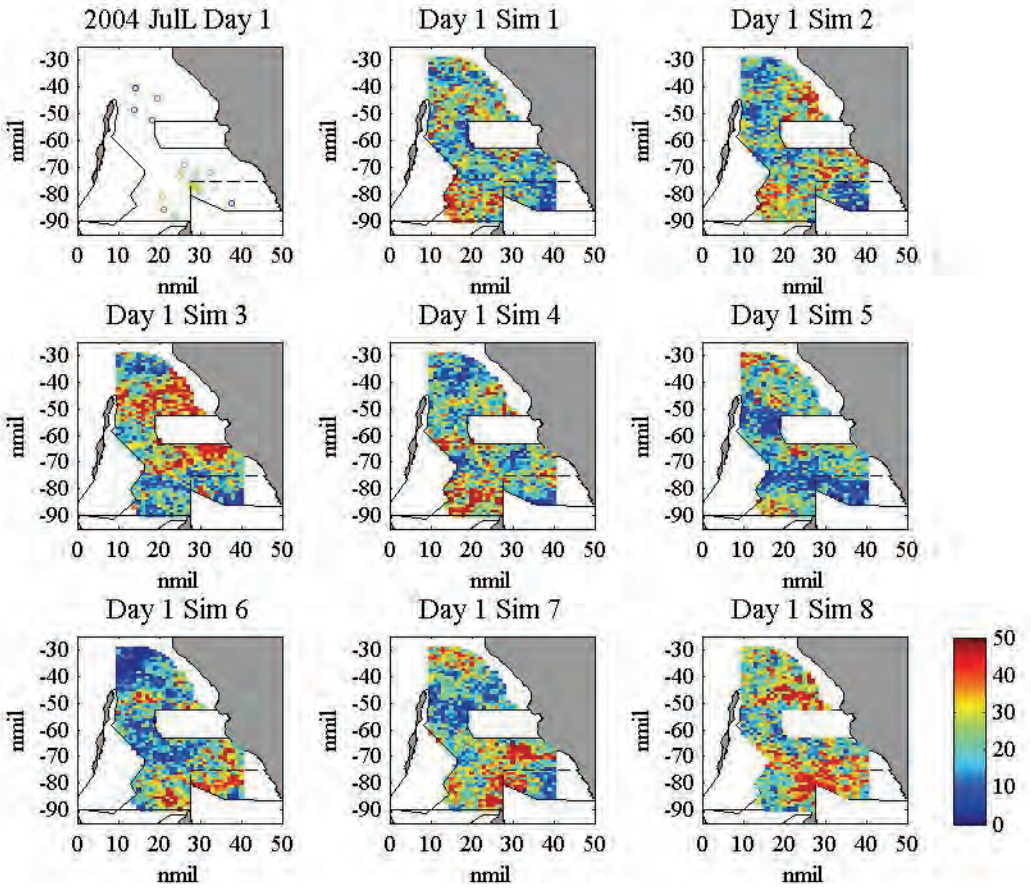


Figure 9.3. *CPUE* values of JulL 2004 (top left) and 8 simulations of *CPUE* values for Day 1 of JulL 2004.

Whilst there are many differences between the separate realisations, a common feature of all realisations is a region of high values to the western side of the ENA. This region of high values is also evident in the actual data for all days of JulL (Figure 9.4) but with a lower maximum. A feature which occurs in a large proportion of the realisation sets has a higher certainty. So the simulated values have correctly identified the area of high values next to the ENA. High values in this region are expected as they are evident in the June data used as input in the simulation process (Figure 9.5).

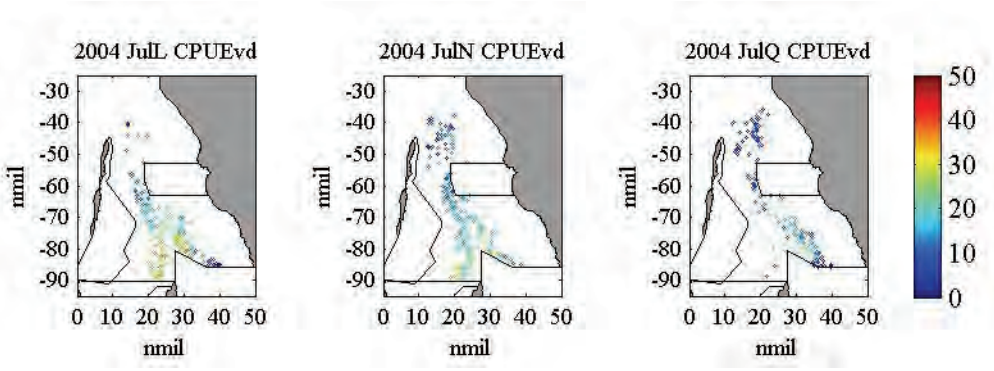


Figure 9.4. *CPUE* values of lunar weeks of July 2004 to be simulated.

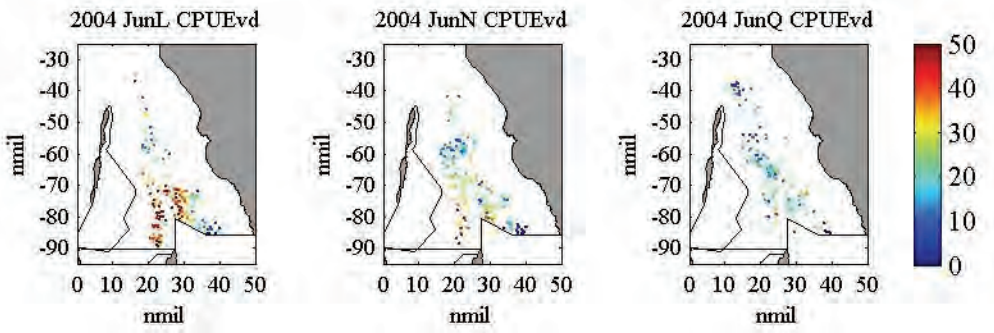


Figure 9.5. *CPUE* values of lunar weeks of previous month (June) 2004 used as conditioning data for simulation of *CPUE* values in July 2004.

The realisations of simulated catch rate values for the JulN week (Figure 9.6) show that, as for JulL, the simulated catch rates overestimate the high catch rates next to the ENA region. Plots of the simulated catch rate values for JulQ (Figure 9.7) indicate few areas of high catch rate values in any of the simulated maps. The locations of these high catch rates also vary over the realisations, suggesting that high catch rate values are not likely during this week. This matches the actual catch rate data of the week of JulQ.

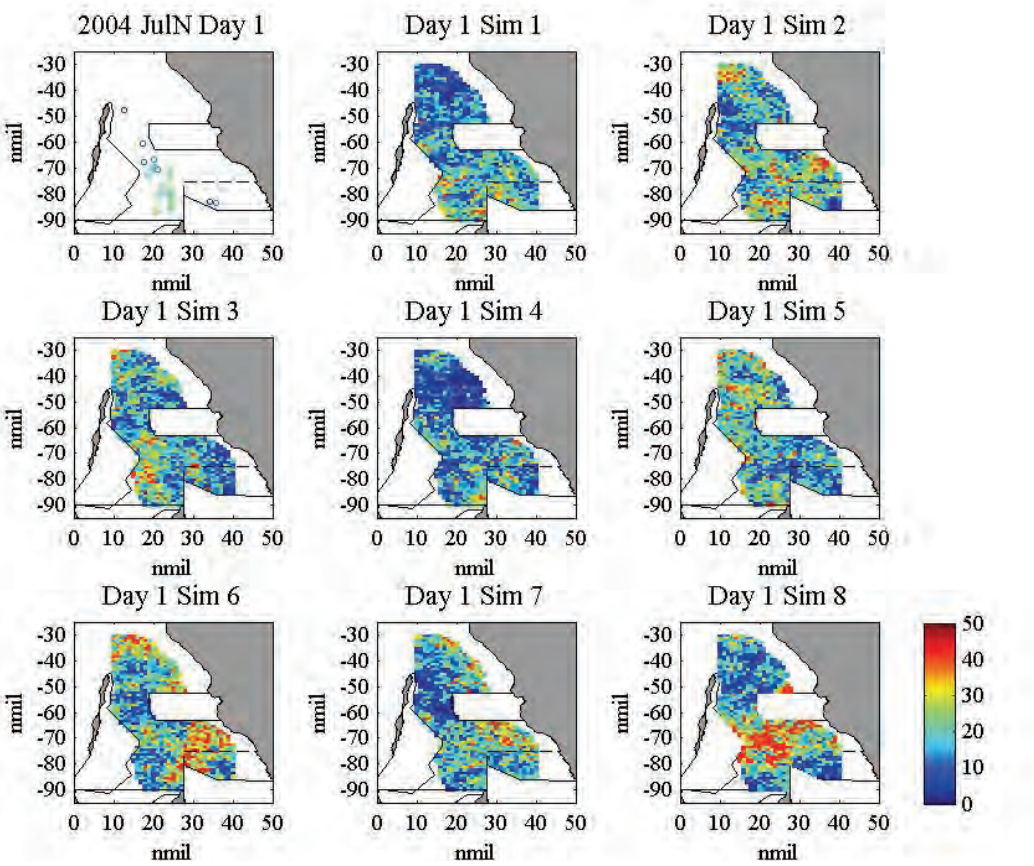


Figure 9.6. *CPUE* values of JulN 2004 (top left) and 8 simulations of *CPUE* values for Day 1 of JulN 2004.

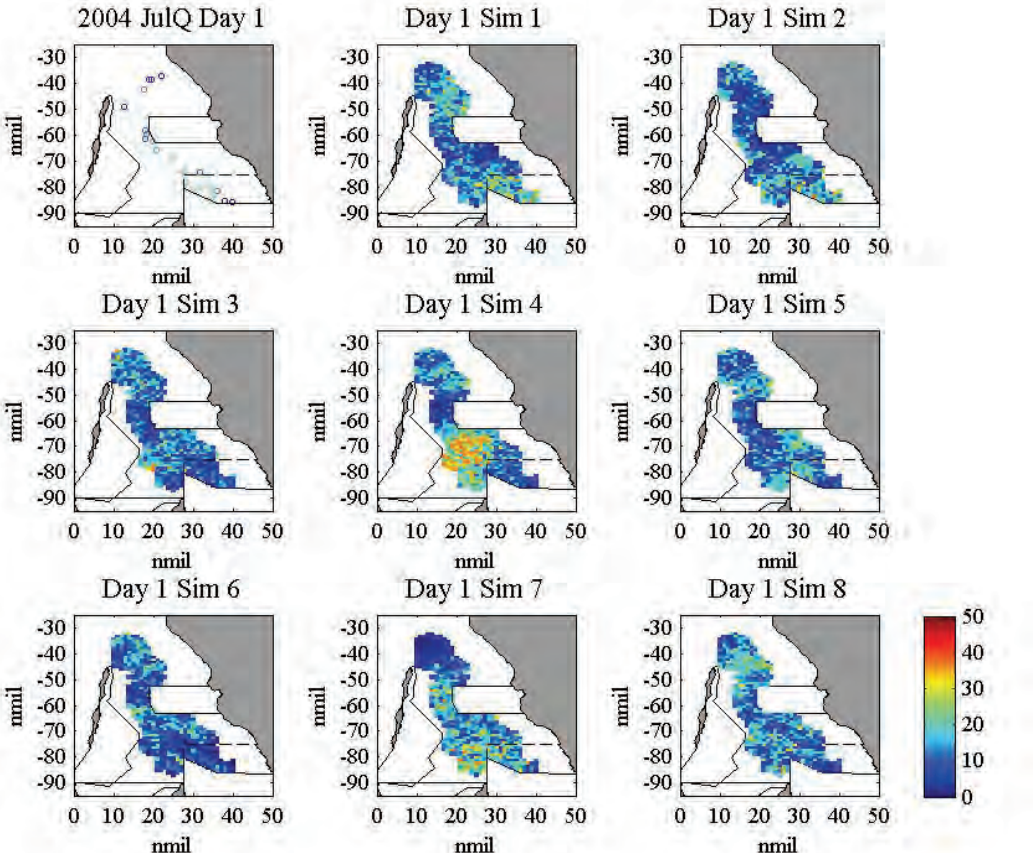


Figure 9.7. *CPUE* values of JulQ 2004 (top left) and 8 simulations of *CPUE* values for Day 1 of JulQ 2004.

9.3.2 SIMULATION OF MAY 2004 CATCH RATE VALUES

The simulation process was also used to compute daily realisations of the catch rates on a 1 nmil by 1 nmil grid for the May lunar month using input data from April (Figure 9.8). It should be noted that the estimates for May failed to produce estimates as far into the ENA as evident in the actual May sample values as May marked the opening of that region for fishing. Furthermore, the estimates that were made in the ENA failed to capture the high values evident in the sample area. Additional (survey) data was used in the estimation process but was not used in the simulation process.

As was the case for the July data the first simulation of the first day from the three weeks of May 2004 (Figure 9.9, second row) shows the decrease in the number of locations being simulated at in the successive weeks. Also similar to the July simulations, the mean simulation of the first day of each week (third row) across the region show even more smoothed maps than the corresponding kriging estimates (fourth row). These mean simulations associate low catch rates with the northern part of the region and a region of intermediate values on the western side of the ENA region, most notably in MayL.

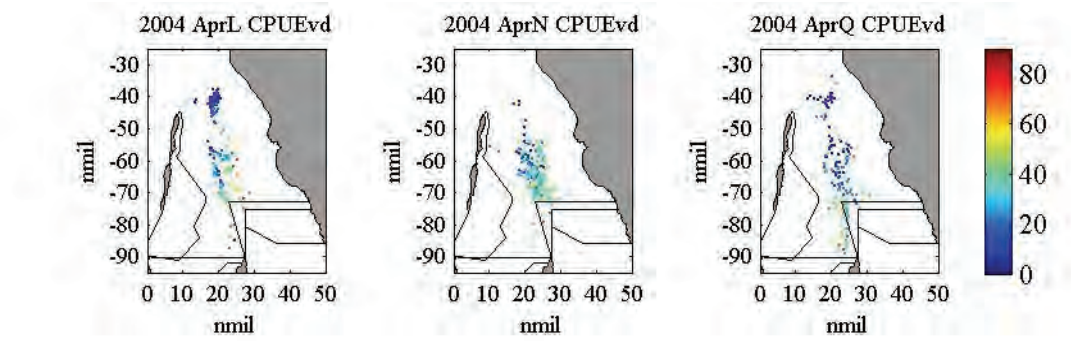


Figure 9.8. *CPUE* values of lunar weeks of previous month (April) 2004 used as conditioning data for simulation of *CPUE* values in May 2004.

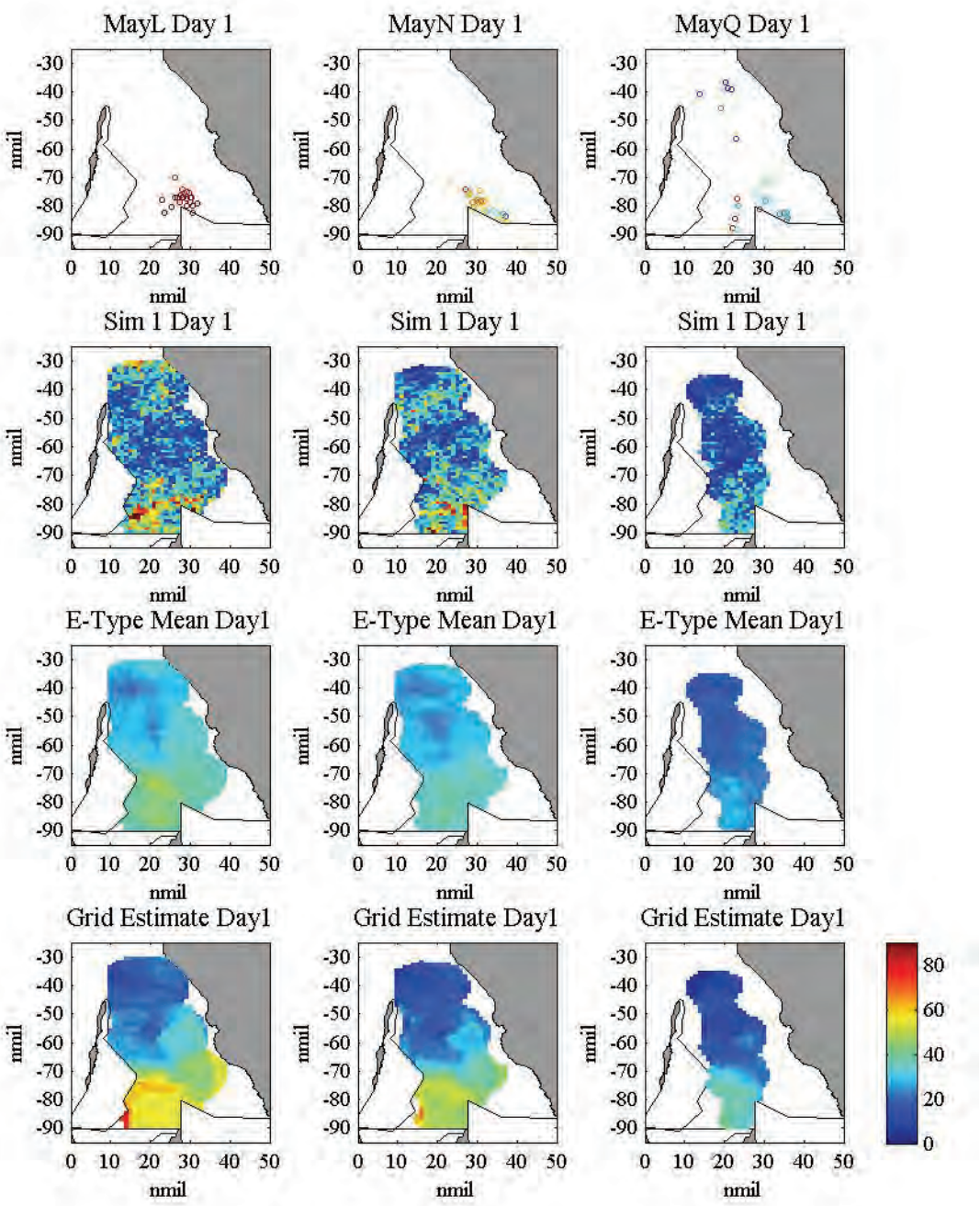


Figure 9.9. *CPUE* values (top row), individual simulation of *CPUE* values (second row), mean of 100 simulations of *CPUE* values (third row) and OK grid estimates of *CPUE* values for the first day of the three lunar weeks of May 2004: MayL (left), MayN (centre) and MayQ (right).

Looking at multiple realisations for the first day of the MayL lunar week (Figure 9.10), the simulated values do not appear to have captured the high catch rates evident in the ENA where the fishers have focussed their efforts, indicated by black rectangle, as it is opened to fishing. This is as expected as there are no data for this area in the April input data (Figure 9.8). Most of the realisations show various small regions of high values spread across the region, which can be regarded as uncertain.

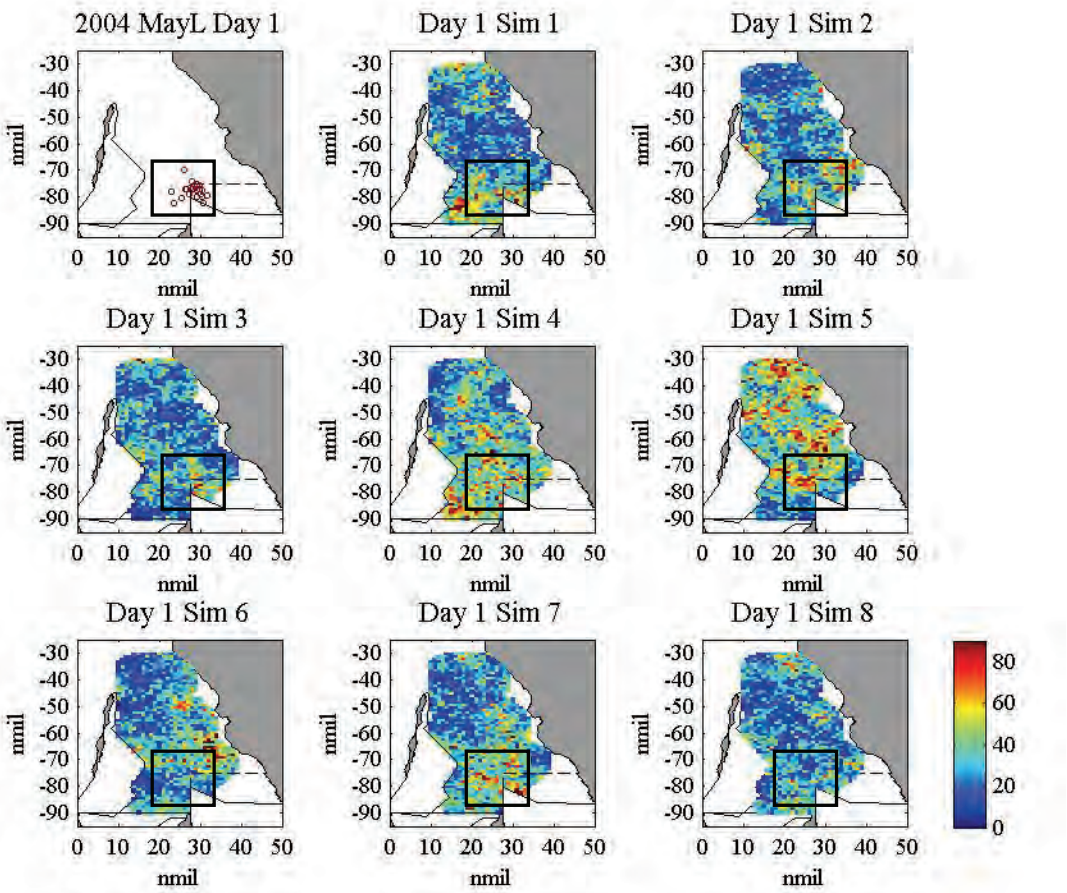


Figure 9.10. *CPUE* values of MayL 2004 (top left) and 8 simulations of *CPUE* values for Day 1 of MayL 2004.

The first eight realisations of simulated catch rate values for the first day of MayN (Figure 9.11) show that, as for MayL, the simulated catch rates do not extend as far into the ENA as the actual catch rate values and they do not capture the high catch rates evident there. The simulated values of the first day of MayQ reflect the mainly low values scattered around the region but little evidence of the few high catch rates evident in the sample data on the western edge of the ENA.

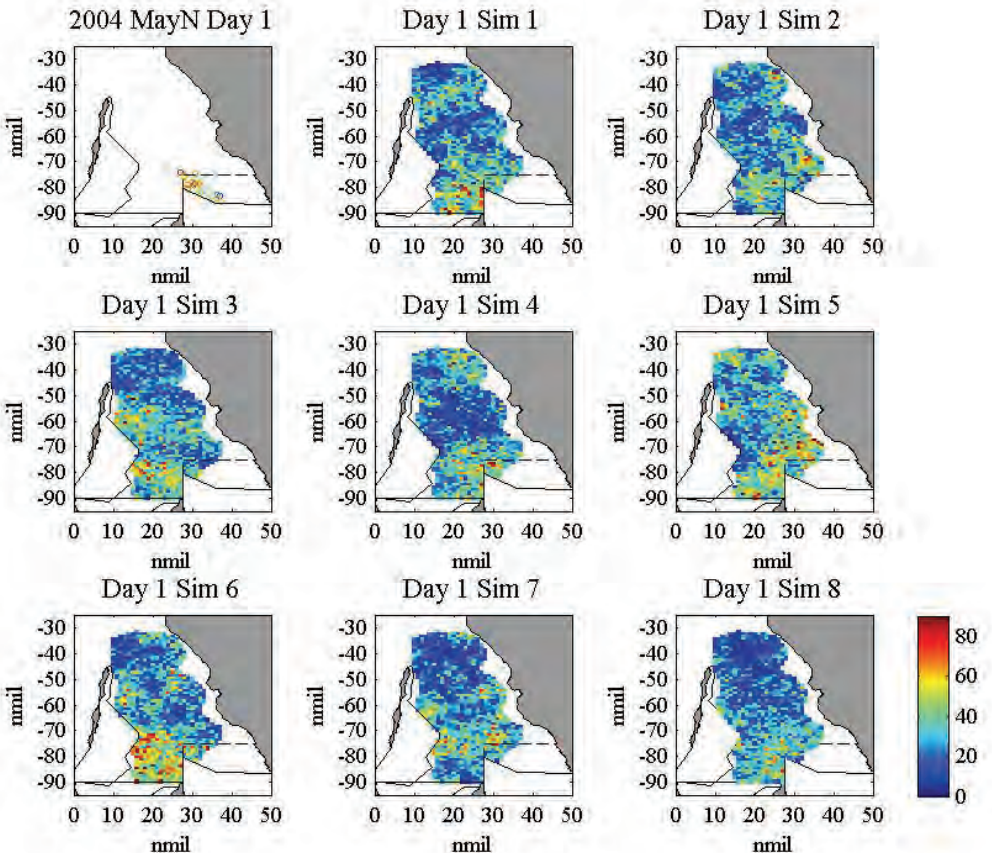


Figure 9.11. Simulated CPUE values, MayN 2004, Day 1, 8 simulations.

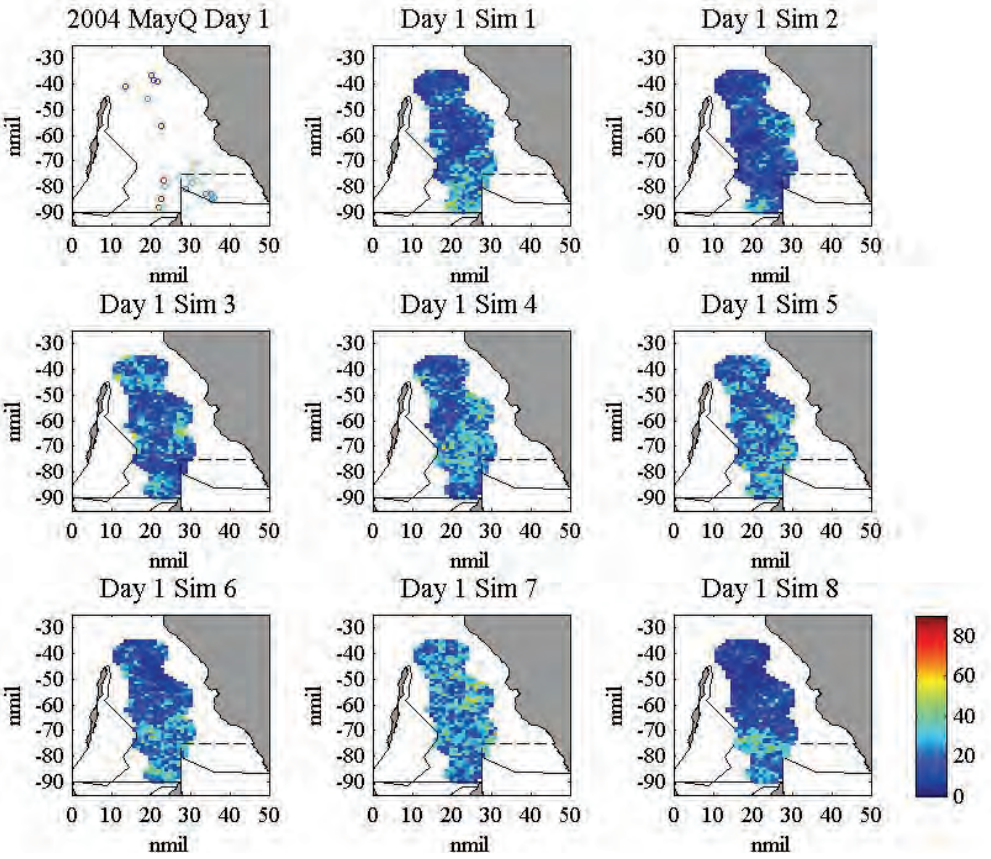


Figure 9.12. Simulated CPUE values, MayQ 2004, Day 1, 8 simulations.

9.4 LOCAL VARIABILITY OF SIMULATIONS

This section focuses on selected grid locations within the Shark Bay region to investigate the local variability at these locations via simulation. The 14 grid locations shown in Figure 9.13 were chosen to cover the fishing region of 2004. Estimates made previously at these grid locations, along with the first individual simulation and the E-type mean simulation for the first day of each lunar week of July are shown in Figure 9.15. As noted previously, the E-type means of the 100 simulations (third row) identify the same areas of high and low values but produce a more smoothed map than that of the estimates (top row), whilst the individual simulations (by design) show much greater variability. Further investigation of this variability is seen in the individual simulations.

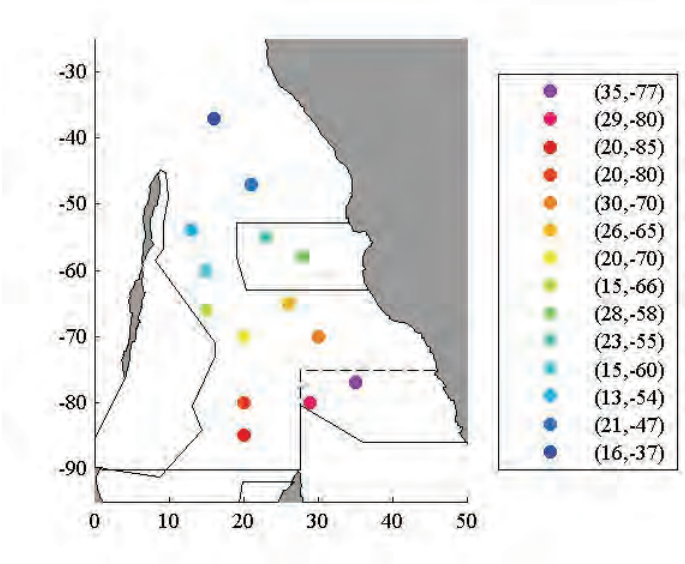


Figure 9.13. Selected locations in fishing region chosen to investigate local variability of *CPUE* simulations over the fishing region

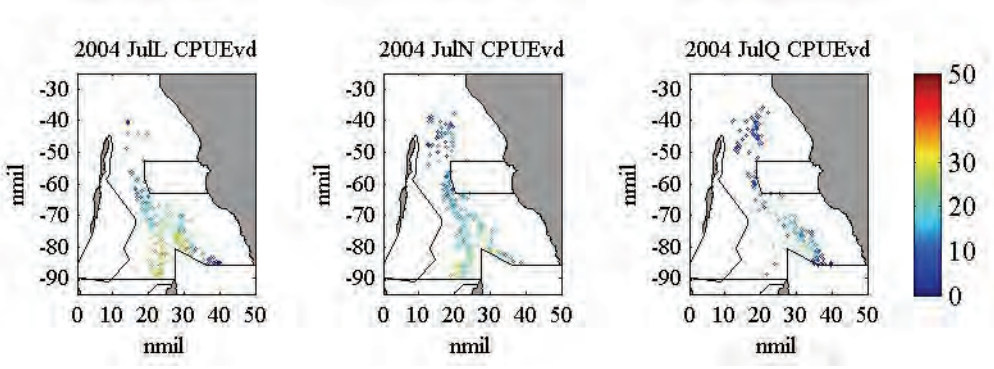


Figure 9.14. *CPUE* values of the three lunar weeks of July 2004 to be simulated.

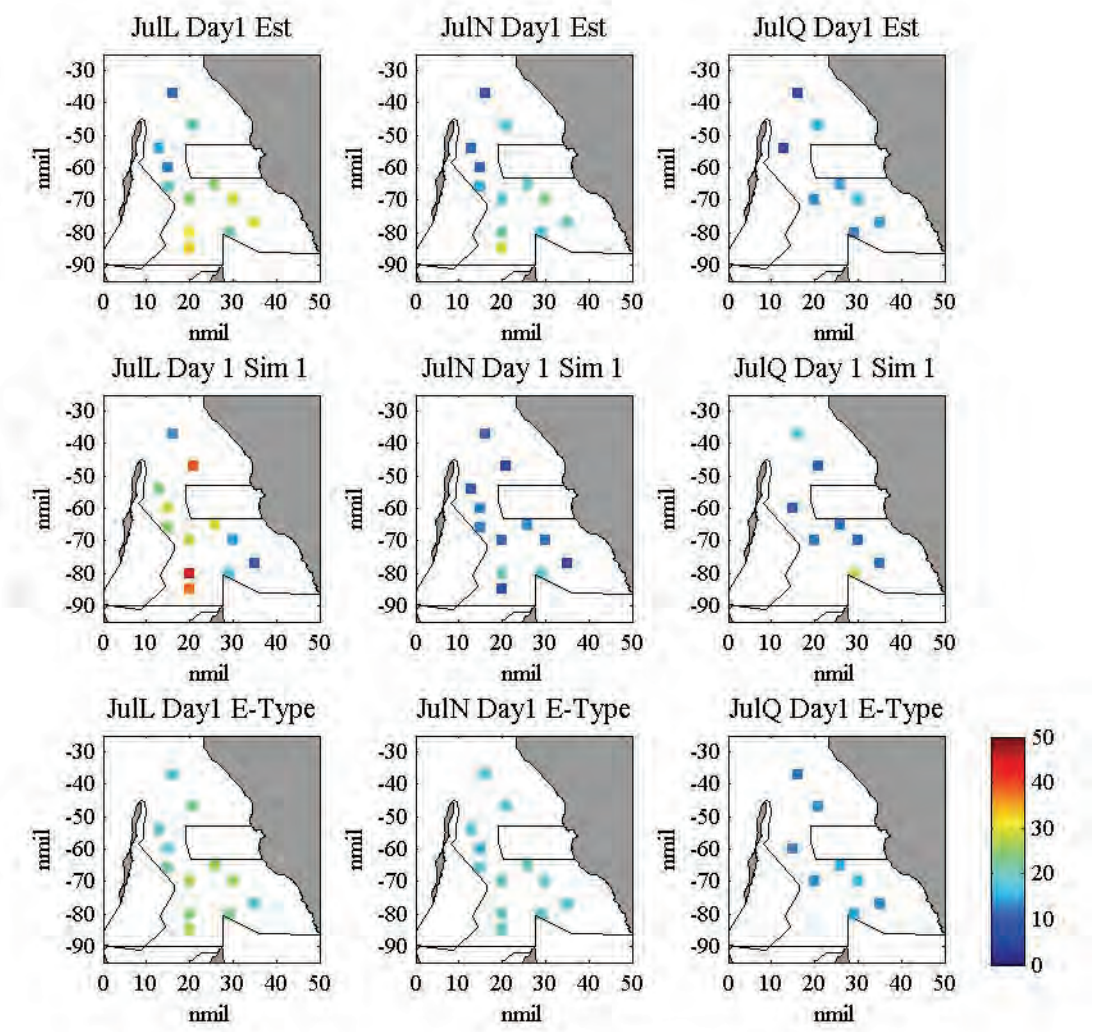


Figure 9.15. *CPUE* OK grid estimates (top row), individual *CPUE* simulation (middle row) and mean of 100 *CPUE* simulations (bottom row) at selected locations, for the first day of each of the three lunar weeks of July 2004.

Cumulative distribution function (cdf) curves were produced at each location for the 100 simulation values at that location for the first day of each lunar week of July (Figure 9.16). The cdfs show the cumulative probability levels associated with varying catch rate values. It is possible to infer the probability level associated with a certain catch rate value or the catch rate associated with a certain cumulative probability level. There is more variability with the cdfs of Day 1 JulL than with those of JulN or JulQ. There is a general pattern of behaviour with the cdfs of the varying locations. In all 3 weeks, the cdfs associated with locations in the lower half of the region (red and yellow) lie to the right of the cdfs associated with locations in the upper half of the region (blue). The cdfs of the locations in the ENA (pink/purple) appear to lie in the middle. This indicates overall higher values for the southern region and lower values for the northern region. From these observations we can conclude that the catch rates in the lower half of the region are associated with higher values than those in the upper region.

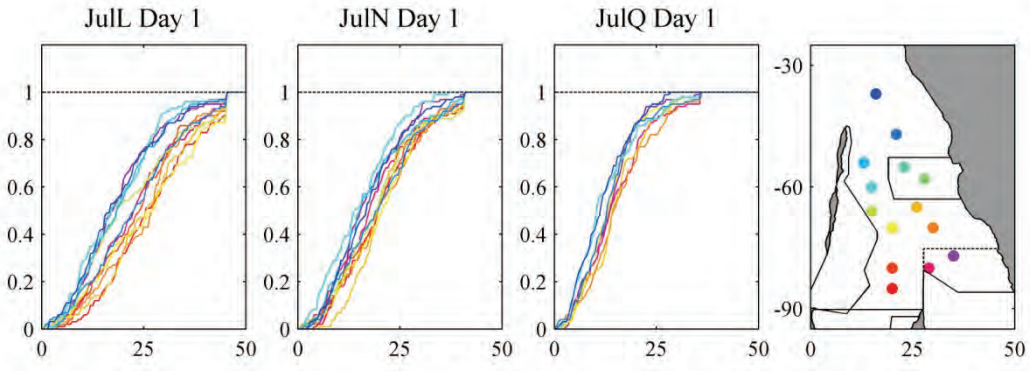


Figure 9.16. Cdf swarms of *CPUE* simulated values at selected locations for the first day of the three lunar weeks of July 2004. Selected locations indicated on map at right.

Boxplots for the 100 simulations at each location (Figure 9.17) confirm these observations of higher and lower values. The boxplots indicate the spread of simulated values at each location. The interquartile range, or IQR, (indicated by the height of the box) shows that the spread of values is greater for the central locations in JulL (yellow) and lowest for all locations in JulQ. This suggests that the central region in July may be the area of most uncertainty in July.

A comparison of the estimates obtained at each of the selected locations via kriging with the simulated values indicate that on day 1 of JulL, estimates for the locations just in or near the ENA (pink/red/orange) are near the upper quartile of the corresponding simulated values. The actual catch rate values near these locations tend towards the estimated value but do show a spread of values as indicated by the simulated values. Central locations (yellow) are near the median and locations in the north region (blue) show values near the lower quartile. There is only one data value near these locations so it is difficult to determine its accuracy. The estimates of JulN show similar behaviour to JunL with the actual 2004 catch rate data displaying similar values to those estimated. The range of values indicated by the simulations capture the actual catch rate data values. The estimates of JulQ all lie near the median of the simulated values with the actual catch rate values displaying a range of values across those indicated by simulation. So, the kriging estimates in the south in the lunar month of July are in the upper range indicated by simulation, which is reflected in the actual catch rates seen at these locations in July 2004. Estimates in the centre region are similar to the average simulation and also reflect the actual catch rate values. Estimates in the south are in the lower range of the simulations and are indicative of the catch rates at those locations in July 2004.

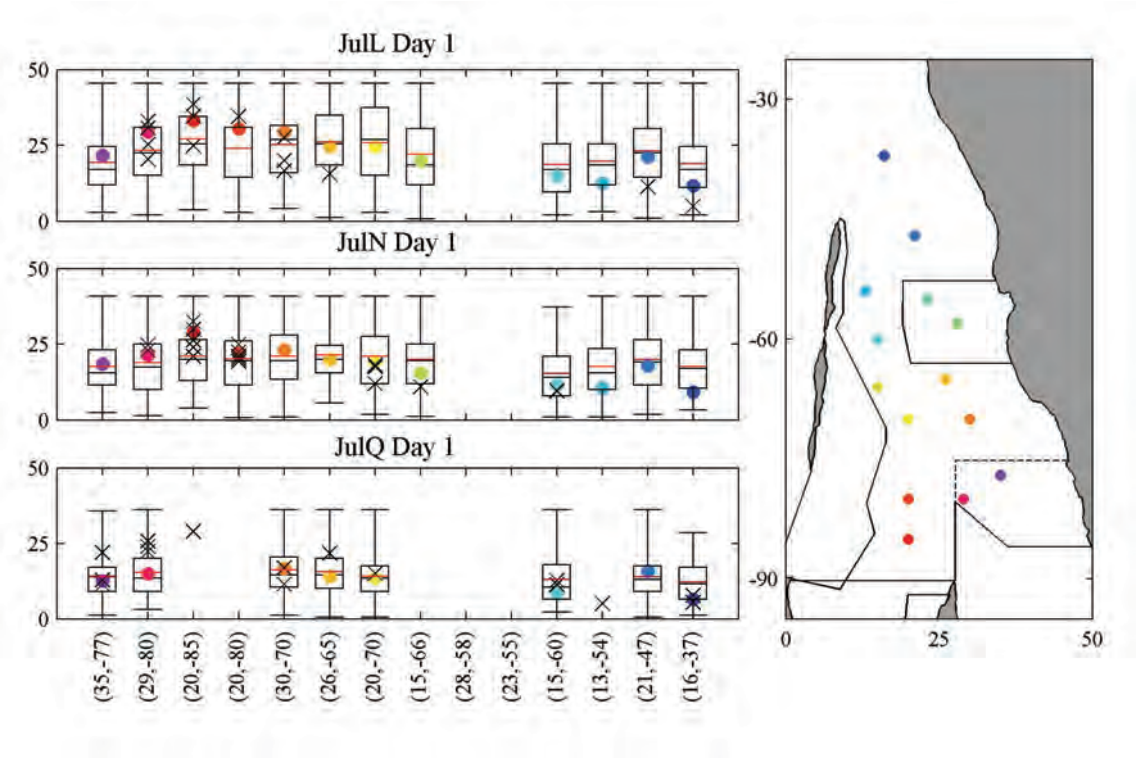


Figure 9.17. Boxplots of *CPUE* simulations with average simulation value (red line) and estimated value (coloured dot) at selected locations for first day of each of the three lunar weeks of July 2004. Black crosses indicate actual catch rate values from selected day of 2004 season within a 5 nmil radius of the location. Selected locations indicated on map at right.

The same 14 locations (Figure 9.13) were used to investigate the local variability at these locations in the lunar month of May. Estimates made previously at these grid locations, along with the first individual simulation and the E-type mean simulation for the first day of each lunar week of May (Figure 9.18) show that the E-type means of the 100 simulations (third row) identify the same areas of high and low values but produce a more smoothed map than that of the estimates (top row). As expected, the individual simulation shows much greater variability.

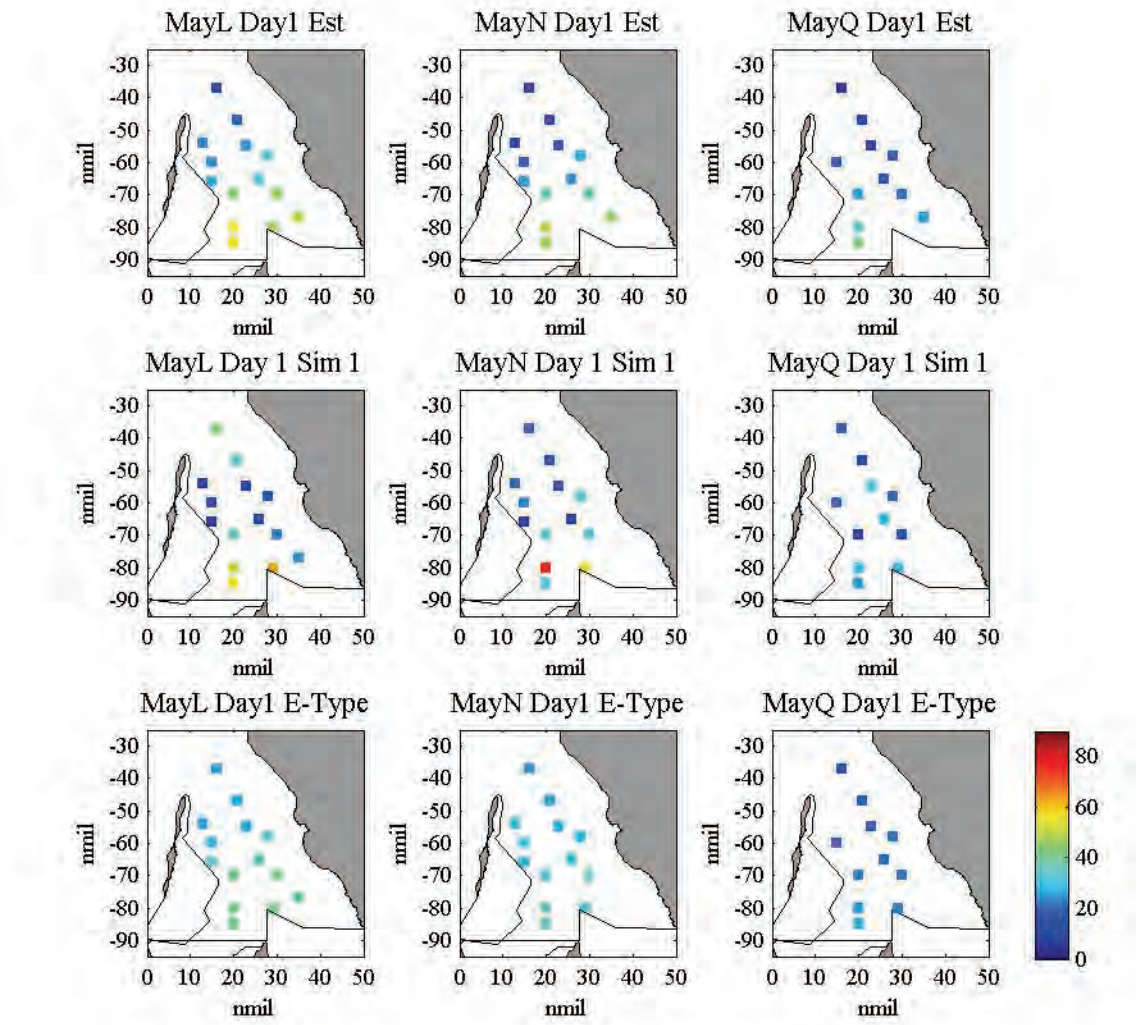


Figure 9.18. *CPUE* OK grid estimates (top row), individual *CPUE* simulation (middle row) and mean of 100 *CPUE* simulations (bottom row) at selected locations, for the first day of each of the three lunar weeks of May 2004.

The cdf swarms for the first day of each lunar week of May at these locations (Figure 9.19) show similarities across the weeks. The locations of the first week show more variability than the other two weeks, which was also seen in the July cdfs. Also similar to July, the cdfs associated with locations in the southern half of the region (red and yellow) lie to the right of the cdfs associated with locations in the upper half of the region (blue). This demonstrates that the catch rates in the southern half of the region are associated with higher values than those in the northern region. The shapes of the cdfs for May are quite different to those of July, as they show a much longer tail to their upper limit indicating a more strongly positively skewed distribution.

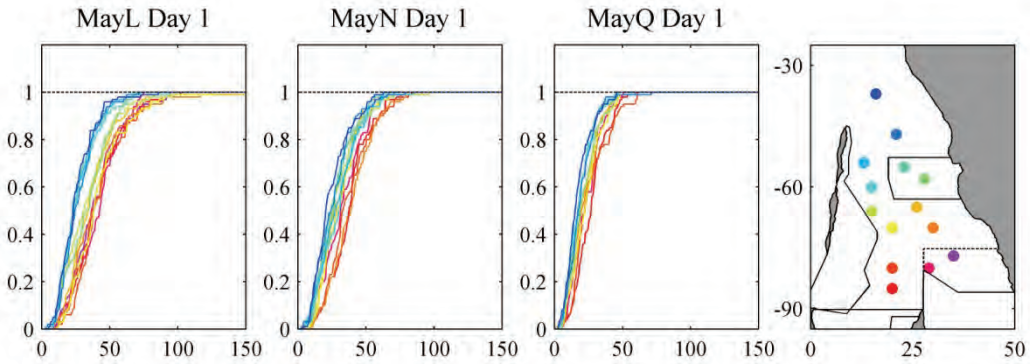


Figure 9.19. Cdf swarms of *CPUE* simulated values at selected locations for the first day of the three lunar weeks of May 2004. Selected locations indicated on map at right.

Boxplots for the 100 simulations for the first day of each lunar week in May (Figure 9.20) show a much greater variability in the spread of values at the various locations than seen in the July simulations. The locations in the centre and southern half of the region (red/yellow) have a greater spread in values than those in the northern part of the region (blue). This suggests that the lower values in the northern region during May are of more certainty. The central and southern regions hold much more uncertainty. The boxplots also confirm the regions of higher and lower values.

Across all weeks of May, kriging estimates for the locations just in or near the ENA (pink/red/orange) are near the upper quartile of the corresponding simulated values and are reflected in the actual catch rates seen in 2004. Estimates at the central locations (yellow) are near the median and locations in the northern region (blue) show values near the lower quartile. There are few actual catch rates near these locations to verify these values, but of those within 5 nmil of the estimated locations, they show values in the range indicated by the simulated values.

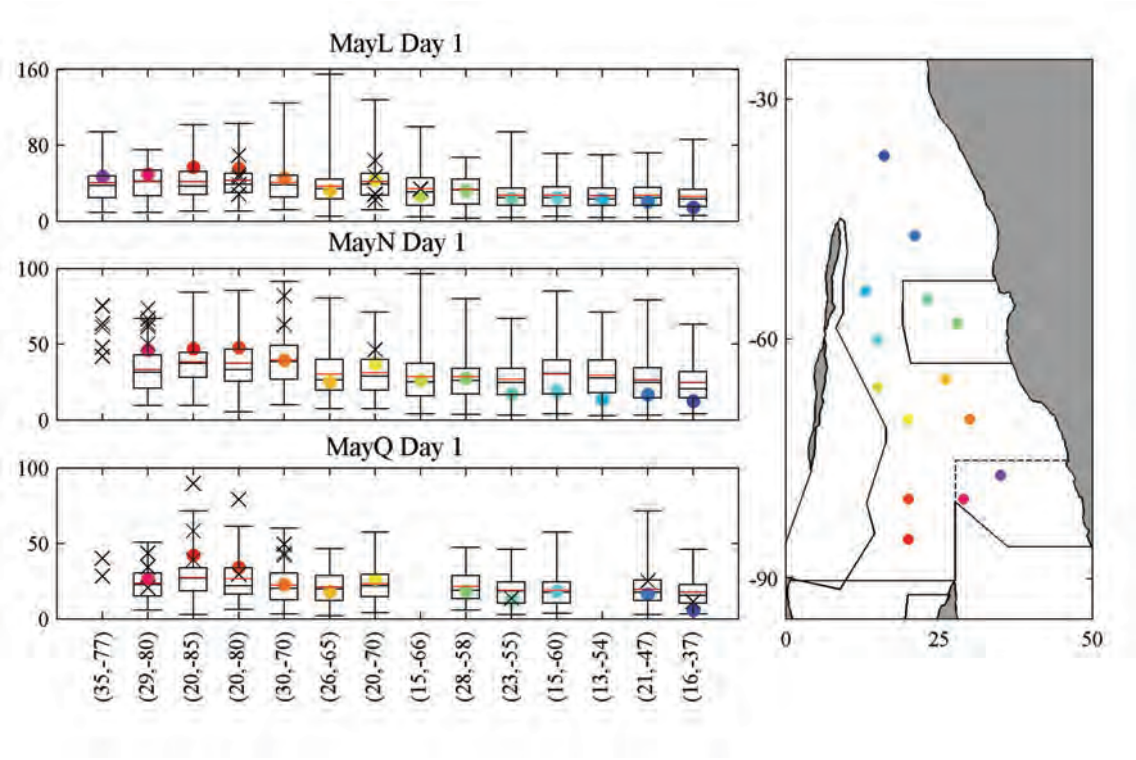


Figure 9.20. Boxplots of *CPUE* simulations with average simulation value (red line) and estimated value (coloured dot) at selected locations for lunar weeks of May 2004. Black crosses indicate actual catch rate values from selected day of 2004 season within a 5 nmil radius of the location. Selected locations indicated on map at right.

9.5 PROPORTION OF VALUES ABOVE A THRESHOLD

Using the 100 simulations at each grid location, maps of proportions of the simulated values above a given threshold can be produced for a given threshold. Figure 9.21 and Figure 9.22 show the proportions for the first day of each week of July and May respectively where the catch rate is greater than the global trend model for that temporal period. This is equivalent to the probability that the standardised catch rate value is greater than one, and therefore indicates areas where the catch rate is higher than the average value.

For the northern region in both May and July weeks, the probability that the simulated catch rate will be greater than the global trend is at most 0.5 in all weeks. Therefore the northern region will most likely involve catch rate values that are well below the temporal trend value for that week. There is an area on the western side of the ENA and another area above the ENA show proportions between 0.25 and 0.5 in both lunar months, confirming regions of low values (below the temporal trend) suggested by kriging estimates and seen in the actual data. In May there is a small region on the western side of the ENA showing proportions greater than 0.5, most notably in MayL. This confirms the areas of high catch rate values in this region. The region of high values evident in the actual catch rates in the ENA in MayL, and to a lesser extent, during MayN is not captured in the simulation proportion maps of May.

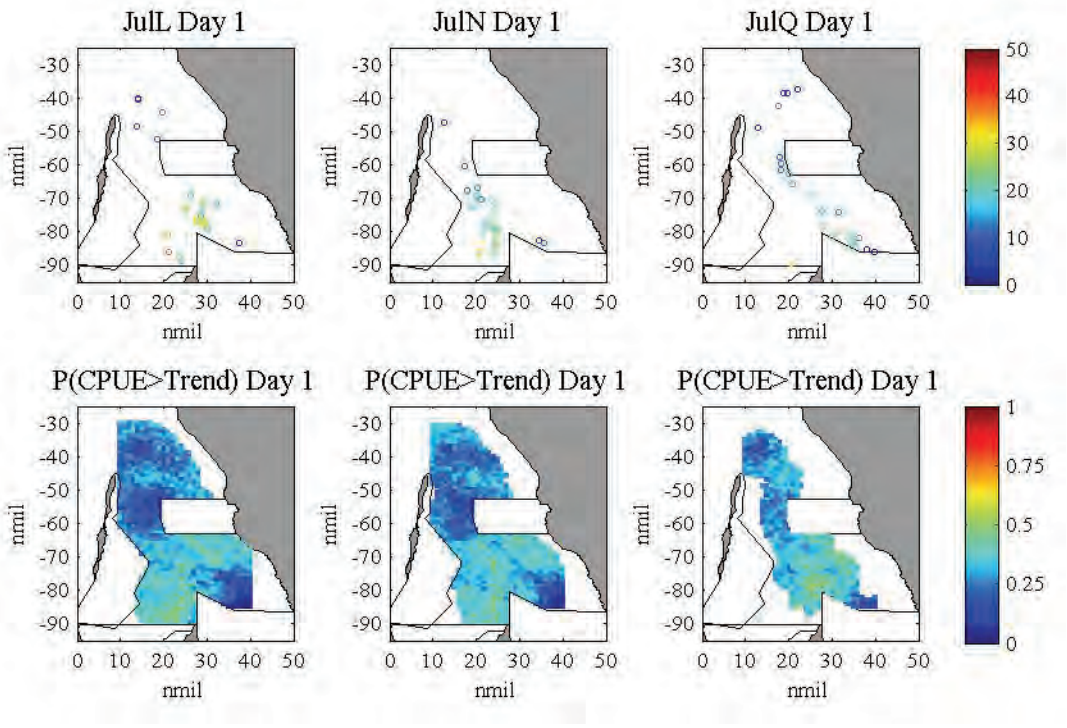


Figure 9.21. *CPUE* values (top row) and proportion of simulated values at each location whose value exceeds the temporal trend model (bottom row) for the lunar weeks of July 2004.

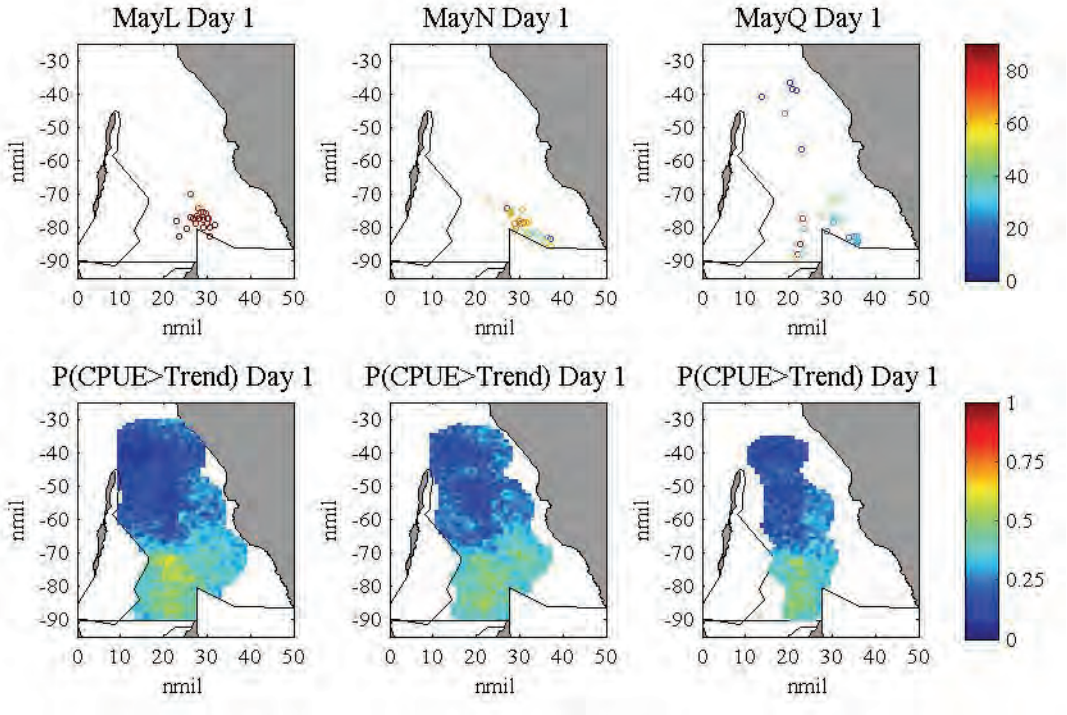


Figure 9.22. *CPUE* values (top row) and proportion of simulated values at each location whose value exceeds the temporal trend model (bottom row) for the lunar weeks of July 2004.

9.6 CHAPTER SUMMARY

Simulation of the king prawn catch rates provided another way to not only deduce estimates over the spatiotemporal domain, but also to investigate the local variability of the values. In fact, the optimal estimate is derived from the model of local uncertainty. As for the estimation

process, the simulation process could produce alternative catch rate realisations for the subsequent lunar month of fishing. Simulated values identified areas of higher or lower uncertainty across the spatiotemporal domain, by suggesting greater or less variability at each location. However, average simulation values were smoother than the kriging estimates, which may be related to the search parameters of the kriging procedure or the cdf extrapolation/interpolation parameters. The ability of the simulation process to identify the probability of the variable being above or below a given threshold value could be utilised in the fishery to identify possible areas of concern that could be used to direct management decisions. It should be noted that the deterministic trend used to obtain the standardised catch rate reduced the space of uncertainty. The uncertainty measures in this chapter come only from the space-time variability of the residuals and do not account for the variability of the temporal trend.

PART IV.

GENERAL CONCLUSIONS AND FURTHER RESEARCH

CHAPTER 10

GENERAL CONCLUSIONS AND FURTHER RESEARCH

10.1 INTRODUCTION

This thesis had three main objectives: (i) To develop a geostatistical spatiotemporal modelling framework for the Shark Bay prawn catch rate data based on the single spatiotemporal random function model that will model and allow future predictions of the catch rate along with estimates of uncertainty; (ii) To apply the developed geostatistical space-time model to the Shark Bay prawn catch rate data using the data from the 2001 to 2003 seasons to compute estimates/simulations for the 2004 season which will then be compared with the actual 2004 values; and (iii) To integrate fishery-independent survey data to improve the performance of estimates obtained for the Shark Bay prawn catch data from the developed geostatistical space-time model. This chapter summarises the major outcomes of this thesis and their contribution to the accomplishment of these objectives. Recommendations for further research are also given.

10.2 DISCUSSION

Spatiotemporal geostatistical modelling is an empirical approach to the modelling of the space-time distribution of a variable, with its parameters estimated from observed data. Although not physically based, a space-time geostatistical model must include the differences in its structure and parameterisation to address the fundamental differences between spatial and temporal variation (Kyriakidis & Journel, 1999a; Heuvelink & Griffith, 2010). Space represents a state of coexistence, in which there can be multiple dimensions (or directions), and interpolation is usually important. Conversely, time represents a state of successive existence, where a clear ordering is present that is non-reversible and in only one dimension. In temporal analysis, extrapolation is usually important. The origins of spatial and temporal variation in the variable of interest can be quite different (Snepvangers et al., 2003).

Spatiotemporal geostatistical modelling has typically been divided into two main areas. One is focussed on models involving multiple, temporally correlated, spatial random functions (Egbert & Lettenmaier, 1986; Goovaerts & Sonnet, 1993; Papritz & Flühler, 1994; Bogaert & Christakos, 1997) or spatially correlated time series (Solow & Gorelick, 1986; Rouhani & Wackernagel, 1990; Rouhani et al., 1992). These methods resulted from the typical imbalance between the quantities of temporal and spatial information. The second, and by far the most

commonly implemented, method involves a single spatiotemporal random function model (De Iaco, 2002; Gneiting et al., 2005; Skoien & Bloschl, 2005; Ekstrom et al., 2007; Soares & Pereira, 2007; Gething et al., 2008; Denham & Mueller, 2010). This latter method forms the basis of the spatiotemporal model implemented in this thesis.

Applications of space-time geostatistics, and in particular the single spatiotemporal random function modelling approach, have been growing over the past few decades, albeit restricted mainly to academic research. Practical implementations of the single space-time random function model are data-driven (like most geostatistical methods) and typically face the same dilemmas. The main obstacles are the choice of global trend modelling or transformation method to provide a stationary spatiotemporal variable and the choice of spatiotemporal covariance or semivariogram model that is used to model the spatiotemporal continuity of the stationary variable.

The identification of spatiotemporal global trends is more complicated than for the individual spatial or temporal cases, due to the interaction of space and time. Typically spatial or temporal analysis of the variable has preceded the spatiotemporal analysis, and may have identified a dominant spatial or temporal trend. However, the interaction of space and time must be considered when detrending, and is typically realised by analysis of the residuals in space and time from the chosen global trend model.

One of the main concerns with modelling spatiotemporal correlation structures, as with spatial correlation modelling, is to ensure a valid model. This must satisfy the negative semi-definiteness condition for variogram models or positive-definiteness for covariance functions. A well-established set of models exists for spatial-only variograms (Goovaerts, 1997), while the list for spatiotemporal models is more diverse (De Iaco, 2010). These include, but are not limited to, the product model (Rodríguez-Iturbe & Mejía, 1974), the metric model (Dimitrakopoulos & Luo, 1994), the integrated product model (Cressie & Huang, 1999) and the product-sum model (De Cesare et al., 2001b).

The choice of spatiotemporal covariance model is typically driven by the data, but is also constrained by the user's knowledge of possible models and their implementation. Lists of the commonly used spatial covariance/semivariogram models and their visual representations are well-known and documented whilst those of the spatiotemporal models are not as well distinguished. There are numerous texts outlining the choice of spatial models (Isaaks & Srivastava, 1989; Goovaerts, 1997; Deutsch & Journel, 1998), but not for the spatiotemporal

case. A recent study (De Iaco, 2010) outlines the characteristic behaviours of the more commonly implemented spatiotemporal covariance models, and is a useful guide for anyone wishing to implement a geostatistical spatiotemporal covariance analysis.

Spatiotemporal geostatistics offers various techniques to optimise use of sampled measurement information for interpolating or extrapolating variables at unsampled locations and times. Modelling of a variable from survey data forms the basis of most geostatistical studies. Data may be obtained from independent survey data or from monitoring programs already in place. Whilst survey data is usually obtained from a suitable sampling design for the study at hand, data from monitoring programs do not always satisfy the sampling requirements and must be validated for their use in a spatiotemporal geostatistical model. Using data already obtained to enable an increase of insight into the spatiotemporal distribution of a variable without increasing the measurement effort, is of considerable value as sampling and monitoring are often expensive in environmental studies.

The outcomes of a geostatistical model are typically estimates or simulations of the variable of interest using the chosen kriging algorithm, which is a best linear unbiased estimator. Estimation at an unsampled location or time instant via the kriging process provides a least-squares estimate of the attribute. Each estimate also comes with a precision measure in the form of the kriging variance, albeit a measure of the data configuration. Whilst point estimates are useful, it is also beneficial to obtain a measure of the distribution of the variable at that spatiotemporal location. It is possible to achieve estimates of uncertainty through implementation of a geostatistical simulation process (Goovaerts, 1997).

Although there is a wealth of free and commercial software programs for implementing spatial geostatistical methods, the direct implementation of spatiotemporal methods is not possible in all of these software packages. A comprehensive list of geostatistical software and their functionality is available in Goovaerts (2010). GSLIB (Deutsch & Journel, 1998) is a commonly used suite of programs for spatial geostatistics. Modified versions of GSLIB Fortran routines are available for univariate and multivariate space-time semivariogram modelling (using the product or product-sum models) and for the implementation of kriging routines (De Cesare et al., 2002; De Iaco et al., 2010). As the GSLIB code is freely available, users proficient in programming can also modify the relevant code to address the needs of their individual study. A modified version of the GSLIB *sgsim* routine was produced in this study to enable space-time sequential Gaussian simulation. The constant need for modifying code to address the issue at

hand is possibly the biggest stumbling block in the implementation of space-time geostatistical models but necessary for continued development. This study highlights that the capacity of space-time geostatistical methods to become a useful tool for researchers is reliant on the development of more user-friendly software and documentation of a framework of implementation.

The spatiotemporal continuity of the catch rate data used in this study exhibited a global nugget effect not evident in other studies using modified GSLIB routines. A solution was provided to simulate the global nugget effect by recalculating the marginal temporal semivariogram for use in a product-sum model, based on the spatial aggregation evident in the data. This study highlights the need for the inclusion of a global nugget effect in future modifications of GSLIB routines to allow greater flexibility in modelling spatiotemporal semivariograms. Similarly, these routines would benefit from the addition of alternative spatiotemporal semivariogram models.

Although modified versions of GSLIB routines were used and developed in this study to implement spatiotemporal geostatistical modelling, it must be acknowledged that there are other software packages available for model implementation. However, like the GSLIB routines, not all spatiotemporal covariance structures or kriging variants are available in any one program. The R software suite is identified as one of the fastest growing and most comprehensive statistical computing tools, largely due to its being freeware, the increasing number of packages and its growing online communities (Hengl, 2007). The *gstat* and *spacetime* packages are two of the geostatistical packages offered in R, and allow space-time semivariogram computation and modelling (with limited choices), and their implementation in subsequent estimation and simulation routines (Pebesma & Wesseling, 1998; Bivand et al., 2008). Further development of these packages to include various semivariogram models and kriging and simulation variants will be invaluable to the implementation of space-time models in the wider community.

10.2.1 APPLICATION TO SHARK BAY PRAWN CATCH RATE DATA

Analysis of fisheries stocks by spatial geostatistics has been implemented in a number of studies (Addis et al., 2009; Moura & Fernandes, 2009). These typically involve sample data obtained from a fishery-independent survey designed for the study. Whilst it is preferable to obtain data that are independent of a fishery, the implementation of a fishery-independent survey can be impractical and/or costly (Maunder et al., 2006). There is often a range of fishery-dependent

data available for many fisheries that form part of their ongoing monitoring programs. The use of an already existing data source alleviates the need for additional surveys. This study makes use of the fishery-dependent logbook commercial catch and effort data collected for the Shark Bay managed prawn fishery.

Catch per unit effort (CPUE), also known as catch rate, is a widely used fishery-dependent indicator variable, often as a measure of relative abundance, although this can be problematic (Walters, 2003). CPUE trends are an indicator of the general 'health' of the fishery providing useful insight into the changes within a fishery and highlighting areas of possible concern (Maunder et al., 2006). Annual CPUE values are already used in the Shark Bay managed prawn fishery to compare fishing seasons. As catch rate data are available across the spatial range of the fishery and throughout the fishing weeks of each season, it is logical to analyse spatiotemporal continuity within each season and across seasons. This thesis has studied the use of spatiotemporal geostatistical methods on king prawn catch rate data from the Shark Bay managed prawn fishery in Western Australia. A methodological framework for the estimation and simulation of catch rate values within the Shark Bay fishery at future time instants has been presented (Denham & Mueller, 2009, 2010). As CPUE values are computed from catch and effort data that are obtained from fishermen operating within a given fishery, the data does not necessarily conform to guidelines for a sampling variable originating from a specified survey design. However, considering the catch and effort of multiple fishermen by spatial location should remove the obvious biases that would be expected from the non-random search behaviour of fishers (Walters, 2003).

Two methods of aggregation of the daily catch rate shot data were considered in this work. The first involved combining all shots for a given vessel on a given day into a single record by aggregating all catch and effort values at a centroidal location. This was driven by the awareness that a large proportion of the data was already aggregated by vessel due to the recording practices of many skippers. The second type of daily catch rate data was obtained by aggregating the shot data on a 1 by 1 nmil grid over the study region, to remove individual fisher bias. Data sets obtained by the two methods of aggregation for the 2001 to 2004 seasons displayed similar statistical measures. Whilst aggregating by grid may appear to preserve a finer spatial resolution for the data, the large proportion of data already aggregated by vessel for a day may have been the over-riding problem for this method.

The aggregated data, by vessel or grid, were considered as point data for the purposes of geostatistical variography, estimation and simulation. Whilst it is recognised that these data are not strictly “point” data as they have an associated spatial area, it is difficult to calculate the actual support size of the raw data.

In a previous study (Mueller et al., 2008), the spatial aspect of the king prawn catch rates in Shark Bay showed an amenability to geostatistical methods. This thesis has investigated temporal trends of the king prawn catch rates modelled previously to obtain standardised catch rates that also display a pattern of spatial continuity. The cubic temporal trends seen in previous seasons of catch rate data (Harman, 2001) were also present in the 2001-2004 seasons. As the main objective of the modelling framework was to produce a model that allowed future estimates, the 2001-2003 seasons were used to produce a temporal trend model that could be utilised for the 2004 season.

Standardisation is essential, in any observational data set, to allow undistorted comparisons of rates over time or over different areas (Bishop et al., 2008). By standardising the catch rates via a temporal trend, it was possible to compare daily and weekly catch rates within and across seasons in order to construct a spatiotemporal geostatistical model for king prawn catch rates. The persistence of spatial continuity between successive weeks of a season and comparable weeks between seasons justifies the modelling of the spatiotemporal continuity of the catch rate variable.

10.2.1.1 SPATIOTEMPORAL CONTINUITY

Spatiotemporal continuity was evident in the experimental spatiotemporal semivariograms of the standardised king prawn catch rates. As outlined previously, a critical stage in the process of geostatistical spatiotemporal modelling is the choice of model for the spatiotemporal variogram or covariance function and the estimation of the model parameters. The product-sum semivariogram model does not impose constraints of symmetry between spatial and temporal components. Furthermore, it does not require an arbitrary space-time metric, which is often unreasonable or difficult to formulate. It allows intuitive and simple construction of models for the spatiotemporal continuity, which can be fitted to data using relatively straightforward techniques that are similar to those established for spatial-only geostatistics. Finally, but perhaps most importantly, it captured the behaviour of spatiotemporal continuity evident in the catch rate data and therefore was adopted for use in this study with the king prawn catch rates.

Product-sum models were fitted to the experimental semivariograms of the standardised king prawn catch rate aggregated by vessel or grid. Individual seasons and combined successive seasons were analysed via semivariograms to establish a pattern of spatiotemporal continuity over the 2001-2003 seasons. The experimental semivariograms exhibited a global nugget, which was interpreted as mainly a spatial nugget due to the aggregation of catch and effort data by vessel or grid. The two methods of data aggregation resulted in different measures of spatiotemporal continuity, which displayed some similarities and differences between them. Whilst the model parameters varied over the seasons, there was strong evidence of continuity of the model's behaviour to construct a single model for combined seasons. With the objective of modelling the catch rates of the 2004 season, a single model for the combined 2001-2003 seasons was obtained. Ultimately only one method of aggregation (by vessel) was used to conduct estimation on the basis of the fit of a model for spatiotemporal continuity. The performance of the space-time semivariogram model was cross-validated via kriging for the 2004 sample data, one week at a time.

10.2.1.2 SPATIOTEMPORAL ESTIMATION

It is important to be able to map and accurately forecast the location and the spatial characteristics of a fisheries resource for the purpose of stock conservation and profit optimisation (Maynou et al., 1998). The catch rate of the Shark Bay managed prawn fishery has been used for several years as a general indicator of the health of the fishery (Sporer et al., 2010). The continuity described by the spatiotemporal semivariogram model of the standardised catch rate was used in a spatiotemporal geostatistical model to provide estimates at unsampled spatiotemporal locations. Estimates of the standardised catch rate were transformed to catch rate values using the temporal trend model initially used to standardise the catch rates.

The model produced allowed geostatistical estimation of the catch rates in interpolation and extrapolation mode. The focus of this study was extrapolation mode by producing estimates for the fishery in future weeks of the 2004 season. In extrapolation mode, the model could predict the behaviour of the catch rates outside the temporal and spatial bounds of the given data (Denham & Mueller, 2009). The extrapolation process in this thesis was focussed on forecasting, where the catch rate was estimated into the future of the fishing season using data up to, but not including, the month being estimated. In practice, this meant that only data from up to 30 days before the day being estimated were actually used in the estimation process.

The model could reasonably predict approximately two weeks into the future of the 2004 season as long as the input data covered the region to be estimated. This meant that the week corresponding to the first quarter moon phase of each lunar month could not be forecast accurately as it was the third week of the month being estimated. The inclusion of data from the first week of a lunar month increased the ability to make successful estimates of the catch rate in the third week. It also enabled slightly more locations to be estimated in the second week of the month and significantly more in the third week. Furthermore, the extra data provided a slight refinement of the areas of high values already estimated by the original input data.

As new areas were opened to fishing (such as the ENA in the first lunar week of May), there were insufficient data to enable estimation in that region. To be more precise, some estimation was possible in that region, but only on the edge near the region already open. Inclusion of data from the first week of a lunar month proved useful again, not only to extend the temporal range of the estimation process but to provide data within the opening region that was not covered by the original input data used for estimation. However, data in the first week of May of the ENA yielded values much higher than was evident in successive weeks so, whilst this allowed estimates to be made further into the ENA, the estimates for that region in successive weeks were overly biased by the high values evident in the first week as the region was opened to fishing. This was not of concern when using the additional first week of data for the month of July 2004 as there was no opening region or areas of significantly high or low catch rates. A possible solution for weeks/areas involving the opening of a fishing region is to limit the range of influence of high values in a new area as it is opened for fishing by scaling their values for use in estimating future weeks.

In the case of May 2004 which involved an opening region (ENA) in the first week of fishing, survey data available in the ENA prior to fishing was used to increase the number of estimates within that region, and for them to be valid estimates. They provided the input data for that region to be estimated, as well the high values necessary to replicate the high values evident in the area (Denham & Mueller, 2010). The sample data were given an appropriate pseudo-date to ensure their inclusion in the estimation process. The survey data were not useful for estimating in the 2nd and 3rd weeks as they bias the values in that area. Alternatives may be to scale the survey data to reduce their impact in successive weeks, or to use catch rate data from the ENA in previous years to use as input data for a region as it opens. Either way, it is

necessary to use additional input data for use in weeks involving an opening region such as the ENA and for successive weeks until the impact of that opening region has stabilised.

The use of additional data allows more estimates to be computed but must be representative of region. Using data from the first week of a lunar month to predict the 2nd and 3rd weeks of a month will help in refining the estimates if the region has been consistently open (eg. July). However in the case of an opening region involving significant catch rates (eg May) the additional data may enable further estimates, but as they are heavily biased, it is not appropriate. Additional data for a week involving an opening region (MayL) must be representative data (eg. survey or previous year). In the case of survey data, it must be weighted for use in subsequent weeks. In the case of previous year data, only data from the appropriate weeks should be used.

10.2.1.3 SPATIOTEMPORAL SIMULATION

A spatiotemporal simulation process was also implemented with the catch rates as a tool to model the variability of estimates in the lunar months of May and July of 2004. Simulation was conducted using only the original input data so the simulated values within ENA in MayL cannot hope to capture the variability of the catch rates in that area. Only areas outside the ENA in MayL can be checked for variability with current estimates. As the catch rate behaviour is more consistent within the region leading into the month of July, all locations and days in July can be assessed for variability by the simulation process. The simulation process clarified areas of consistently low or consistently high values in these months and identified areas of high variability.

Estimates obtained for the catch rates of the 2004 season via averaging multiple simulated realisations of the spatiotemporal random field provided even smoother results than the kriging estimates. This is not generally expected, and indicates the need to refine the simulation process used for the catch rates. This may involve changing the search parameters, cdf extrapolation parameters or possibly even the local model of variability assumed (multiGaussian).

10.3 FUTURE RESEARCH DIRECTIONS

The geostatistical spatiotemporal model presented in this thesis offers a method to model the spatiotemporal continuity of the king prawn catch rate in the Shark Bay managed prawn fishery. It can be used to compute estimates and simulated values of the catch rate within seasons and approximately two weeks into the future of a season. This development of ideas began with an

initial model for the spatiotemporal behaviour of the catch rate and then aimed to improve the modelling of the catch rates by the inclusion of relevant input data. There are more refinements possible for this model.

Four seasons of king prawn catch rates were analysed in this study. Applying this model to additional seasons of king prawn catch rate data would help to refine and justify the parameters within the model. Although this model has been developed for use with the standardised king prawn catch rate variable, the flexibility of the modelling process means it can be applied to the king prawn catch variable or to the catch or catch rate of another species such as the tiger prawn within the Shark Bay fishing region. Other transformations of the catch rates could also be considered such as standardising by fitting generalised linear models, a method not yet implemented in the Shark Bay managed prawn fishery but a commonly used standardisation technique in many fisheries analysis (Punt et al., 2000; Battaile & Quinn Ii, 2004; Maunder & Punt, 2004). The spatiotemporal continuity of the standardised catch rates obtained via a generalised linear model could be analysed in a similar manner as presented in this study and used to produce estimates or simulations of the catch rates.

The neighbourhood used for the inclusion of relevant temporal and spatial information was based on a space-time metric as implemented by the modified *kt3dnew.exe* routine (De Cesare et al., 2002). This program could be further modified to allow direct control over the number of spatial neighbours used in the estimation process for each temporal separation. For example, allowing the 10 closest spatial neighbours for each temporal separation from 1 to 7 days. The ability of the user to specify the number of spatial neighbours used for each temporal window would be a useful addition to this program. This feature has been implemented in the Edinburgh Space-Time Geostatistics Package (Spadavecchia, 2008) which offers a set of programs for implementing space-time estimation (OK, IK, UK) and sequential Gaussian simulation.

The effect of environmental factors on the dynamic nature of the Shark Bay managed prawn fishery should be considered in further studies. Water temperature and rainfall are two possible factors for consideration. Geostatistical methods are well suited to the inclusion of secondary information, either through deterministic trend modelling of the original variable (Spadavecchia, 2008), modelling of trend within the kriging process as a function of the secondary variables via kriging with an external drift (KED) or simple kriging with a local mean (SKLM) (Goovaerts, 1997), or through modelling of the multivariate space-time covariance

structure (De Iaco et al., 2010). Integration of environmental variables is also possible in a simulation setting via simulation with local means or co-simulation (Goovaerts, 1997). The inclusion of environmental variables could provide more realistic scenarios and increase the performance of the model's estimates or simulations. By achieving a more robust model, fishery scientists and industry members of the Shark Bay managed prawn fishery would be interested to gain information from forecasting catches in future months, particularly for contrasting environmental conditions (M. Kangas, Personal communication, October 2011).

Additional input data could be used for estimating over the whole region or their use could be limited to within a specific region. Where an opening region is involved it may be best to use the additional data only for estimating within that region. For example, the actual data from MayL indicates that there is a distinct line of behaviour from outside to inside the ENA, so it is appropriate to limit the use of additional data (survey data or previous year's data) to estimating within the ENA. This would mean estimates for MayL would be a combination of estimates using original data and survey data within ENA with the original estimates for the remainder of the region. Given that the behaviour in the opening region ENA is so vastly different to the areas near the region in previous weeks, it may be even better to estimate within the ENA using only input data from within the ENA (ie survey data or previous year/s data). Estimates outside the ENA would be made using only data outside the ENA, which essentially is just the original data used for estimation.

The estimation and simulation procedures within this modelling framework utilised the linear methods of OK and SK. Whilst optimal for normally distributed data (Johnston et al., 2001), these methods can also work with non-normally distributed data. The skewed nature of the catch rate data may also lend itself to the use of non-linear kriging algorithms, which are actually linear kriging (SK or OK) applied to specific non-linear transforms of the original data. Non-linear kriging methods, such as Lognormal Kriging (LK) and Indicator Kriging (IK), are used extensively in spatial geostatistics (Maravelias et al., 1996; Fernandes & Rivoirard, 1999; Kishné et al., 2003; Kleisner et al., 2010a).

As catch rate data are often modelled by a lognormal distribution, LK is an obvious alternative to the ordinary kriging process employed in this study (Goovaerts, 1997), although studies are limited to spatial geostatistical methods. The experimental logarithmic semivariogram was investigated in this study but did not show improvement over the traditional semivariogram and therefore not considered further. However, an alternative standardisation

procedure may suggest a logarithmic distribution is appropriate. LK works on kriging the transformed data using the semivariogram of the transformed values and then back-transforming the estimates into the original scale of data. Back-transformation is achieved as the exponential of the kriging estimate plus a nonbias term. By taking the skewed distribution of the data into account, the transformation aims to reduce the data variance to improve the calculation of statistics and weighted averages such as kriging estimates than that which would have been obtained when it is ignored. However, the lognormal hypothesis is very strict so any departure results in a biased expected value of the estimate and the nonbias term is completely dependent on the semivariogram model. It has been suggested in a spatial study that the use of an appropriate factor to correct estimates in the logarithmic domain can eliminate the biasedness of the back-transformed LK estimates (Yamamoto, 2007).

Although IK is another non-linear approach for skewed data, it is not recommended here for use with the prawn catch rates due to the series of threshold values or cut-offs required between the smallest and largest data values in the set. IK builds the cdf at each point based on the semivariogram/covariance structure of indicator transforms of the data. This requires the computation and modelling of semivariograms for each threshold which can be cumbersome and time-consuming. Simulation offers a more viable alternative to the modelling of the cdf. The simulation process employed in this thesis was based on a multiGaussian framework, where kriging was applied to the normal score transforms of the data. By refining the method of standardisation for the catch rates, the performance of this simulation method may be improved.

Uncertainty measures were presented for the space-time variability of the standardised catch rates but underestimated the global uncertainty due to the choice of a deterministic temporal trend model. Future research could also include quantification of the variability of the data around the temporal trend model to provide a more accurate representation of the global uncertainty.

A method of quality control for geostatistical simulations is computation of the spatiotemporal semivariograms of the simulated values. Such variograms can also serve as validation of the simulation model and are recommended for future presentation of this work.

Whilst the model was used for forward prediction in this study, it can also be used in an interpolation mode which aims to describe the spatiotemporal behaviour of the catch rates within the spatial and temporal bounds of the data. Estimates produced for all spatial locations within the fishery for all days within a fishing season, can be used as an analysis tool for the

season that has passed to help refine management approaches. The inclusion of the temporal variable may be able to improve the accuracy of the spatial estimation maps obtained previously (Mueller et al., 2008). Spatial maps are seen as a useful tool for communication between scientific staff and the wider community, including stakeholders.

This thesis has presented a geostatistical spatiotemporal modelling framework for the king prawn catch rates in the Shark Bay managed prawn fishery off Western Australia. Data from other fisheries could also be used in a similar modelling context. There is wide scope to refine and improve the model for use with the catch rates or another fishing variable. With increasing ability in the modern world to readily collect detailed spatial and temporal information of fishery-dependent variables, spatiotemporal modelling of fisheries data has a range of possibilities for implementation. A further challenge lies in the integration of the spatiotemporal modelling potential of the geostatistical framework into the fisheries management regime. It is only with the wider and more frequent use of models and simulations such as those presented in this thesis, that this will gradually be achieved through demonstrating the relevance of the outputs to assessing stock status and evaluating future harvest strategies.

REFERENCES

Adams, C. F., Harris, B. P., & Stokesbury, K. D. E. (2008). Geostatistical comparison of two independent video surveys of sea scallop abundance in the Elephant Trunk Closed Area, USA. *ICES Journal of Marine Science: Journal du Conseil, 65*(6), 995-1003.

Addis, P., Secci, M., Manunza, A., Corrias, S., Niffoi, A., & Cau, A. (2009). A geostatistical approach for the stock assessment of the edible sea urchin, Paracentrotus lividus, in four coastal zones of Southern and West Sardinia (SW Italy, Mediterranean Sea). *Fisheries Research, 100*(3), 215-221.

Armstrong, M. (1998). *Basic Linear Geostatistics*. New York: Springer-Verlag.

Armstrong, M., Chetboun, G., & Hubert, P. (1993). Kriging the rainfall in Lesotho. In A. Soares (Ed.), *Geostatistics Troia '92* (Vol. 2, pp. 661-672). Dordrecht, Boston: Kluwer Academic Publishers.

Battaile, B. C., & Quinn Ii, T. J. (2004). Catch per unit effort standardization of the eastern Bering Sea walleye pollock (Theragra chalcogramma) fleet. *Fisheries Research, 70*(2-3), 161-177.

Bevilacqua, M., Mateu, J., Porcu, E., Zhang, H., & Zini, A. (2010). Weighted composite likelihood-based tests for space-time separability of covariance functions. *Statistics and Computing, 20*(3), 283-293.

Bilonick, R. A. (1985). The space-time distribution of sulfate deposition in the northeastern United States. *Atmospheric Environment, 19*(11), 1829-1845.

Bishop, J., Venables, W. N., Dichmont, C. M., & Sterling, D. J. (2008). Standardizing catch rates: is logbook information by itself enough? *Journal of Marine Science, 65*, 255-266.

Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. New York: Springer.

Bogaert, P., & Christakos, G. (1997). Stochastic analysis of spatiotemporal solute content measurements using a regression model. *Stochastic Hydrology and Hydraulics, 11*(4), 267-295.

Bruno, F. (2004). *Non-separability in space-time covariance functions.* (Doctoral dissertation). Universita degli Studi di Bologna, Bologna.

Bruno, F., & Cocchi, D. (2004). Non-separability in spatiotemporal covariance models arisen by non-stationarity in time. *Atti della XLII Riunione Scientifica della SIS-Bari 2004, CLEUP Padova*, 157-160.

Bruno, F., Guttorp, P., Sampson, P. D., & Cocchi, D. (2009). A simple non-separable, non-stationary spatiotemporal model for ozone. *Environmental and Ecological Statistics, 16*(4), 515-529.

Campbell, R. A. (2004). CPUE standardisation and the construction of indices of stock abundance in a spatially varying fishery using general linear models. *Fisheries Research, 70*(2-3), 209-227.

Castrejón, H., Pérez-Castañeda, R., & Defeo, O. (2005). Spatial structure and bathymetric patterns of penaeoid shrimps in the southwestern Gulf of Mexico. *Fisheries Research, 72*(2-3), 291-300.

Chatfield, C. (2004). *The Analysis of Time Series* (6th Edition ed.). Florida: Chapman and Hall/CRC.

Chiles, J.-P., & Delfiner, P. (1999). *Geostatistics: Modeling spatial uncertainty.* New York: John Wiley & Sons, Inc.

Christakos, G. (2000). *Modern Spatiotemporal Geostatistics*. Oxford: Oxford University Press.

Christakos, G., & Hristopulos, D. T. (1998). *Spatiotemporal Environment Health Modelling*. Boston: Kluwer Academic Publishers.

Colloca, F., Bartolino, V., Lasinio, G. J., Maiorano, L., Sartor, P., & Ardizzone, G. (2009). Identifying fish nurseries using density and persistence measures. *Marine Ecology Progress Series, 381*, 287-296.

Cressie, N. (1993). *Statistics for spatial data* (Revised ed.). New York: Wiley.

Cressie, N., & Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association, 94*(448), 1330-1340.

De Cesare, L., Myers, D. E., & Posa, D. (1997). Spatial-temporal modeling of SO_2 in Milan district. In A. Soares, J. Gomez-Hernandez & R. Froidevaux (Eds.), *GeoENV I: Geostatistics for environmental applications: Proceedings of the Geostatistics for Environmental Applications Workshop, Lisbon, Portugal, 18-19 November 1996.* Dordrecht, Boston: Kluwer Academic Publishers.

De Cesare, L., Myers, D. E., & Posa, D. (2001a). Estimating and modeling space-time correlation structures. *Statistics and Probability Letters, 51*, 9-14.

De Cesare, L., Myers, D. E., & Posa, D. (2001b). Product-sum covariance for space-time modeling: an environmental application. *Environmetrics, 12*, 11-23.

De Cesare, L., Myers, D. E., & Posa, D. (2002). FORTRAN programs for space-time modeling. *Computers and Geosciences, 28*, 205-212.

De Iaco, S. (2002). *A space-time multivariate analysis for environmental data.* Paper presented at the International Association for Mathematical Geology, 2002, Berlin, Germany.

De Iaco, S. (2010). Space-time correlation analysis: A comparative study. *Journal of Applied Statistics, 37*(6), 1027-1041.

De Iaco, S., Myers, D. E., Palma, M., & Posa, D. (2010). FORTRAN programs for space-time multivariate modeling and prediction. *Computers and Geosciences, 36*, 636-646.

De Iaco, S., Myers, D. E., & Posa, D. (2001). Space-time analysis using a general product-sum model. *Statistics and Probability Letters, 52*, 21-28.

De Iaco, S., Myers, D. E., & Posa, D. (2002a). Nonseparable space-time covariance models: Some parametric families. *Mathematical Geology, 34*(1), 23-42.

De Iaco, S., Myers, D. E., & Posa, D. (2002b). Space-time variograms and a functional form for total air pollution measurements. *Computational Statistics and Data Analysis, 41*(2), 311-328.

Denham, A., & Mueller, U. (2009). Space-time geostatistical analysis of King Prawn catch rate. In J. M. Ortiz & X. Emery (Eds.), *GEOSTATS 2008 – Proceedings of the Eighth International Geostatistics Congress*: Mining Engineering Department, University of Chile.

Denham, A., & Mueller, U. (2010). Incorporating survey data to improve space-time geostatistical analysis of King Prawn catch rate. In P. M. Atkinson & C. D. Lloyd (Eds.), *geoENV VII – Geostatistics for Environmental Applications*. Dordrecht: Springer.

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and User's Guide* (2nd ed.). New York: Oxford University Press.

Dimitrakopoulos, R., & Luo, X. (1994). Spatiotemporal modelling: Covariances and ordinary kriging systems *Geostatistics for the next century* (pp. 88-93).

Egbert, G. D., & Lettenmaier, D. P. (1986). Stochastic Modeling of the Space-Time Structure of Atmospheric Chemical Deposition. *Water Resour. Res., 22*(2), 165-179.

Ekstrom, M., Kyriakidis, P. C., Chappell, A., & Jones, P. D. (2007). Spatiotemporal stochastic simulation of monthly rainfall patterns in the United Kingdom. *Journal of Climate, 20*(16), 4194-4209.

Elmatzoglou, I. (2006). *Spatio-Temporal Geostatistical Models, with an Application in Fish Stock.* (Master's thesis). Lancaster University, Lancaster.

Faraj, A., & Bez, N. (2007). Spatial considerations for the Dakhla stock of Octopus vulgaris: indicators, patterns, and fisheries interactions. *ICES Journal of Marine Science, 64*, 1820-1828.

Fernandes, P. G., & Rivoirard, J. (1999). A geostatistical analysis of the spatial distribution and abundance of cod, haddock and whiting in North Scotland. In J. Gomez-Hernandez, A. Soares & R. Froidevaux (Eds.), *GeoENV II: Geostatistics for environmental applications: Proceedings of the Second European Conference on Geostatistics for Environmental Applications held in Valencia, Spain, November 18-20, 1998.* Dordrecht, Boston: Kluwer Academic Publishers.

Fernandez-Casal, R., Gonzalez-Manteiga, W., & Febrero-Bande, M. (2003). Flexible spatio-temporal stationary variogram models. *Statistics and Computing, 13*, 127-136.

Figueira, R., Sousa, A. J., Pacheco, A. M. G., & Catarino, F. (1999). Space-time geostatistical modelling: A case study of sea-salt measured on lichens. In J. Gomez-Hernandez, A. Soares & R. Froidevaux (Eds.), *GeoENV II: Geostatistics for environmental applications: Proceedings of the Second European Conference on Geostatistics for Environmental Applications held in Valencia, Spain, November 18-20, 1998.* Dordrecht, Boston: Kluwer Academic Publishers.

Fletcher, W. J., & Sumner, N. R. (1999). Spatial distribution of sardine (Sardinops sagax) eggs and larvae: an application of geostatistics and resampling to survey data. *Canadian Journal of Fisheries and Aquatic Sciences, 56*, 907-914.

Fuentes, M. (2006). Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference, 136*, 447-466.

Garcia-Cabrejo, O., Quijano-Nieto, J., Macias-Acevedo, J., Obregon-Neira, N., & Rodriguez-Ordonez, J. (2009). Spatio-temporal modeling of rainfall combining artifical neural networks and geostatistics. In J. M. Ortiz & X. Emery (Eds.), *GEOSTATS 2008 – Proceedings of the Eighth International Geostatistics Congress*: Mining Engineering Department, University of Chile.

Gaynor, P. E., & Kirkpatrick, R. C. (1994). *Introduction to Time-Series Modeling and Forecasting in Business and Economics*. New York: McGraw-Hill Inc.

Gething, P., Noor, A. M., Gikandi, P. W., Hay, S. I., Nixon, M. S., Snow, R. W., et al. (2008). Developing geostatistical space-time models to predict outpatient treatment burdens from incomplete national data. *Geographical Analysis, 40*(2), 167-188.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association, 97*(458), 590-600.

Gneiting, T., Genton, M. G., & Guttorp, P. (2005). *Geostatistical space-time models, stationarity, separability and full symmetry*: Department of Statistics, University of Washington.

Gneiting, T., Genton, M. G., & Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability and full symmetry. In B. Finkenstadt, L. Held & V. Isham (Eds.), *Statistical methods for spatio-temporal systems*: Chapman & Hall/CRC.

Goovaerts, P. (1997). *Geostatistics for natural resources estimation.* New York: Oxford University Press.

Goovaerts, P. (1999). Geostatistics in soil science: stats-of-the-art and perspectives. *Geoderma, 89*, 1-45.

Goovaerts, P., & Sonnet, P. (1993). Study of spatial and temporal variations of hydrogeochemical variables using factorial kriging analysis. In A. Soares (Ed.), *Geostatistics Troia '92* (Vol. 2, pp. 745-756). Dordrecht, Boston: Kluwer Academic Publishers.

Granger, C. W. J., & Newbold, P. (1977). *Forecasting Economic Time Series*. New York: Academic Press.

Haas, T. C. (1995). Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *Journal of the American Statistical Association, 90*(432), 1189-1199.

Hall, N. G., & Watson, R. A. (2000). A spatial representation of the tiger prawn (Penaeus esculentus) fishery in Shark Bay, Western Australia. In D. A. Hancock, D. C. Smith & J. D. Koehn (Eds.), *Fish Movement and Migration. Australian Society for Fish Biology Workshop Proceedings, Bendigo, Victoria, September 1999*. Sydney: Australian Society for Fish Biology.

Harley, S. J., Myers, R. A., & Dunn, A. (2001). Is catch-per-unit-effort proportional to abundance? *Canadian Journal of Fisheries and Aquatic Sciences, 58*(9), 1760-1772.

Harman, T. (2001). *The effect of the moon phase on the daily catch rate of king, tiger and endeavour prawns in the Shark Bay and Exmouth Gulf fisheries.* (Honours Thesis). Edith Cowan University, Perth, Australia.

Haslett, J., & Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Applied Statistics, 38*(1), 1-50.

Hengl, T. (2007). *A Practical Guide to Geostatistical Mapping of Environmental Variables*. Luxembourg: Office for Official Publications of the European Communities.

Heuvelink, G. B. M., & Griffith, D. A. (2010). Space-time geostatistics for geography: A case study of radiation monitoring across parts of Germany. *Geographical Analysis, 42*, 161-179.

Heuvelink, G. B. M., Musters, P., & Pebesma, E. J. (1997). Spatio-temporal kriging of soil water content. In E. Y. Baafi & N. A. Schofield (Eds.), *Geostatistics Wollongong '96*: Kluwer Academic Publishers.

Hilborn, R., & Walters, C. J. (1992). *Quantitative fisheries stock assessment: choice, dynamics, and uncertainty.* New York: Chapman and Hall.

Hoaglin, D., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis.* New York: Wiley.

Hoaglin, D., Mosteller, F., & Tukey, J. W. (1985). *Exploring data tables, trends and shapes.* New York: Wiley.

Hoaglin, D., Mosteller, F., & Tukey, J. W. (1986). *Understanding robust and exploratory data analysis.* New York: Wiley.

Isaaks, E., & Srivastava, R. (1989). *An introduction to applied geostatistics*. New York: Oxford Univeristy Press.

Johnston, K., Ver Hoef, J. M., Krivoruchko, K., & Lucas, N. (2001). *Using ArcGIS Geostatistical Analyst*: ESRI Press.

Jones, R. H., & Zhang, Y. (1997). Models for continuous stationary space-time processes. In T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren & R. D. Wolnger (Eds.), *Modelling longitudal and spatially correlated data* (Vol. 122 of Lecture Notes in Statistics). New-York: Springer-Verlag.

Jost, G., Heuvelink, G. B. M., & Papritz, A. (2005). Analysing the space-time distribution of soil water storage of a forest ecosystem using spatio-temporal kriging. *Geoderma, 128*, 258-273.

Journel, A., & Huigbregts, C. (1978). *Mining Geostatistics*. London: Academic Press.
Kangas, M., McCrea, J., Fletcher, W., Sporer, E., & Weir, V. (2006). Shark Bay Prawn Fishery. *ESD Report Series No. 3, January 2006*.
Kishné, A. S., Bringmark, E., Bringmark, L., & Alriksson, A. (2003). Comparison of Ordinary and Lognormal Kriging on Skewed Data of Total Cadmium in Forest Soils of Sweden. *Environmental Monitoring and Assessment, 84*(3), 243-263.
Kleisner, K. M., Walter, J. F., Diamond, S. L., & Die, D. J. (2010a). Modelling the spatial autocorrelation of pelagic fish abundance. *Marine Ecology Progress Series, 411*, 203-213.
Kleisner, K. M., Walter, J. F. I., Diamond, S. L., & Die, D. J. (2010b). Modeling the spatial autocorrelation of pelagic fish abundance. *Marine Ecology Progress Series, 411*, 203-213.
Kolovos, A., Christakos, G., Hristopulos, D. T., & Serre, M. L. (2004). Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Advances in Water Resources, 27*, 815-830.
Kyriakidis, P. C. (1999). *Stochastic simulation of spatiotemporal phenomena.* (Doctoral dissertation). Stanford University.
Kyriakidis, P. C., & Journel, A. G. (1999a). Geostatistical space-time models: A review. *Mathematical Geology, 31*(6), 651-684.
Kyriakidis, P. C., & Journel, A. G. (1999b). Stochastic modeling of spatiotemporal distributions: Application to sulphate deposition trends over Europe. In J. Gomez-Hernandez, A. Soares & R. Froidevaux (Eds.), *GeoENV II: Geostatistics for environmental applications: Proceedings of the Second European Conference on Geostatistics for Environmental Applications held in Valencia, Spain, November 18-20, 1998.* Dordrecht, Boston: Kluwer Academic Publishers.
Kyriakidis, P. C., & Journel, A. G. (2001a). Stochastic modelling of atmospheric pollution: a spatial time-series framework. Part 1: Methodology. *Atmospheric Environment, 35*, 2331-2337.
Kyriakidis, P. C., & Journel, A. G. (2001b). Stochastic modelling of atmospheric pollution: a spatial time-series framework. Part II: application to monitoring monthly sulfate deposition over Europe. *Atmospheric Environment, 35*, 2339-2348.
Lophaven, S., Carstensen, J., & Rootzen, H. (2006). Stochastic modelling of dissolved inorganic nitrogen. *Ecological Modelling, 193*(3-4), 467-478.
Ma, C. (2002). Spatio-temporal covariance functions generated by mixtures. *Mathematical Geology, 34*(8), 965-975.
Ma, C. (2003). Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference, 116*, 489-501.
Ma, C. (2005). Linear combinations of space-time covarince functions and variograms. *IEEE Transactions on Signal Processing, 53*(3), 857-864.
Ma, C. (2008). Recent developments on the construction of spatio-temporal covariance models. *Stochastic Environmental Research and Risk Assessment, 22*(Supplement 1), 39-47.
Maravelias, C. D., Reid, D. G., Simmonds, E. J., & Haralabous, J. (1996). Spatial analysis and mapping of acoustic survey data in the presence of high local variability: geostatistical application to North Sea herring (Clupea harengus). *Canadian Journal of Fisheries and Aquatic Sciences, 53*, 1497-1505.
Mateu, J., Porcu, E., & Gregori, P. (2008). Recent advances to model anisotropic space–time data. *Statistical Methods and Applications, 17*, 209-223.
Matheron, G. (1971). *The Theory of Regionalised variables and its applications*. Paris: Les Cahiers du Centre de Morhologie Mathematique de Fontainebleau, Ecole Nationale Superieure des Mines de Paris.

Maunder, M. N., & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research, 70*(2-3), 141-159.

Maunder, M. N., Sibert, J. R., Fonteneau, A., Hampton, J., Kleiber, P., & Harley, S. J. (2006). Interpreting catch per unit effort data to assess the status of individual stocks and communities. *ICES Journal of Marine Science, 63*, 1373-1385.

Maynou, F. X., Sardà, F., & Conan, G. Y. (1998). Assessment of the spatial structure and biomass evaluation of Nephrops norvegicus (L.) populations in the northwestern Mediterranean by geostatistics. *ICES Journal of Marine Science: Journal du Conseil, 55*(1), 102-120.

Mitchell, M. W., Genton, M. G., & Gumpertz, M. L. (2005). Testing for separability of space–time covariances. *Environmetrics, 16*, 819-831.

Moura, O., & Fernandes, A. C. (2009). Kriging density estimates for the bluemouth rockfish, Helicolenus dactylopterus (Scorpaenidae) off the southern Portuguese coast. *Fisheries Research, 95*(1), 112-124.

Mueller, U., Bloom, L., Kangas, M., Cross, J. M., & Denham, A. (2004). Total catch and effort in the Shark Bay King Prawn fishery. In X. Sanchez-Vila, J. Carrera & J. Gomez-Hernandez (Eds.), *geoENV IV: Geostatistics for environmental applications: Proceedings of the Fourth European Conference on Geostatistics for Environmental Applications held in Barcelona, Spain, November 27-29, 2002*. Dordrecht: Kluwer Academic Publishers.

Mueller, U., Kangas, M., Dickson, J., Denham, A., Caputi, N., Bloom, L., et al. (2008). *Spatial and Temporal Distribution of Western King Prawns (penaeus latisulcatus), Brown Tiger Prawns (penaeus esculentus) and Saucer Scallops (amusium balloti) in Shark Bay for Fisheries Management. Project No. 2005/038*. Perth, Western Australia: Fisheries research and Development Corporation, Department of Fisheries, Government of Western Australia and Edith Cowan University.

Murenu, M., Cau, A., Colloca, F., Sartor, P., Fiorentino, F., Garofalo, G., et al. (2010). Mapping the potential locations of European hake (Merluccius merluccius) nurseries in the Italian waters. In T. Nishida & A. E. Caton (Eds.), *GIS/Spatial Analyses in Fishery and Aquatic Sciences* (Vol. 4, pp. 49-68). Saitama, Japan: International Fishery GIS Society.

Myers, D. E. (2002). Space-time correlation models and contaminant plumes. *Environmetrics, 13*, 535-553.

Myers, D. E. (2004). Estimating and Modeling Space-Time Variograms. In H. Todd Mowrer, R. McRoberts & P. C. Van Deusen (Eds.), *Proceedings of the joint meeting of The 6th International Symposium On Spatial Accuracy Assessment In Natural Resources and Environmental Sciences and The 15th Annual Conference of The International Environmetrics Society, June 28 – July 1 2004, Portland, Maine, USA.*

Myers, D. E., & Journel, A. (1990). Variograms with zonal anisotropies and non-invertible kriging systems. *Mathematical Geology, 22*, 779-785.

Papritz, A., & Flühler, H. (1994). Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. *Geoderma, 62*(1-3), 29-43.

Pebesma, E. J., & Wesseling, C. G. (1998). GSTAT, a program for geostatistical modelling, prediction and simulation. *Computers and Geosciences, 24*(1), 17-32.

Penn, J. W., & Caputi, N. (1985). Stock recruitment relationships for the tiger prawn, Penaus esculentus, fishery in Western Australia, and their implications for management. In P. C. Rothlisberg, B. J. Hill & D. J. Staples (Eds.), *Second Australian National Prawn Seminar* (pp. 165-173). Australia: NPS2: Cleveland.

Penn, J. W., Caputi, N., & Hall, N. G. (1995). Spawner-recruit relationships for the tiger prawn, Penaus esculentus, stocks in in Western Australia. *ICES Marine Science Symposia, 199*, 320-333.

Petitgas, P. (1996). Geostatistics and their applications to fisheries survey data. In B. A. Megrey & E. Moksness (Eds.), *Computers in Fisheries Research* (pp. 113-141). London: Chapman and Hall.

Porcu, E., Mateu, J., & Saura, F. (2008). New classes of covariance and spectral density functions for spatio-temporal modelling. *Stochastic Environmental Research and Risk Assessment, 22*(Supplement 1), 65-79.

Posa, D. (1993). A simple description of spatial-temporal process. *Computational Statistics and Data Analysis, 15*, 425-437.

Punt, A. E., Walker, T. I., Taylor, B. L., & Pribac, F. (2000). Standardization of catch and effort data in a spatially-structured shark fishery. *Fisheries Research, 45*(2), 129-145.

Rodrigues, A., & Diggle, P. J. (2010). A Class of Convolution-Based Models for Spatio-Temporal Processes with Non-Separable Covariance Structure. *Scandinavian Journal of Statistics, 37*(4), 553-567.

Rodríguez-Iturbe, I., & Mejía, J. M. (1974). The design of rainfall networks in time and space. *Water Resour. Res., 10*(4), 713-728.

Rouhani, S., Ebrahimpour, M. R., Yaqub, I., & Gianella, E. (1992). Multivariate geostatistical trend detection and network evaluation of space-time acid deposition data--I. Methodology. *Atmospheric Environment. Part A. General Topics, 26*(14), 2603-2614.

Rouhani, S., & Hall, T. J. (1989). Space-time kriging of groundwater data. In M. Armstrong (Ed.), *Geostatistics: Proceedings of the Third International Geostatistics Congress, September 5-9, 1988, Avignon, France* (Vol. 2, pp. 639-650). Dordrecht, Boston: Kluwer Academic Publishers.

Rouhani, S., & Myers, D. E. (1990). Problems in space-time kriging of hydrogeological data. *Mathematical Geology*, 611-623.

Rouhani, S., & Wackernagel, H. (1990). Multivariate geostatistical approach to space-time data analysis. *Water Resources Research, 26*(4), 585-591.

Rufino, M. M., Maynou, F., Abelló, P., Sola, L. G. d., & Yule, A. B. (2005). The effect of methodological options on geostatistical modelling of animal distribution: A case study with Liocarcinus depurator (Crustacea: Brachyura) trawl survey data. *Fisheries Research, 76*(2), 252-265.

Russo, A., Soares, A., Pereira, M. J., & Trigo, R. M. (2009). Geostatistical model for air quality surveillance/monitoring. In J. M. Ortiz & X. Emery (Eds.), *GEOSTATS 2008 – Proceedings of the Eighth International Geostatistics Congress*: Mining Engineering Department, University of Chile.

Sampson, P. D., & Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association, 87*(417), 108-119.

Schabenberger, O., & Gotway, C. A. (2005). *Statistical methods for spatial data analysis*. Boca Raton: Chapman and Hall.

Skoien, J. O., & Bloschl, G. (2005). Spatio-temporal geostatistical analyses of runoff and precipitation. In P. Renard, H. Demougeot-Renard & R. Froidevaux (Eds.), *Geostatistics for Environmental Applications: Proceedings of the Fifth European Conference on goeostatistics for environmental applications.* (pp. on disk). Berlin: Springer-Verlag.

Snepvangers, J. J. J. C., Heuvelink, G. B. M., & Huisman, J. A. (2003). Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma, 112*, 253-271.

Soares, A., & Pereira, M. J. (2007). Space-time modelling of air quality for environmental-risk maps: A case study in South Portugal. *Computers & Geosciences, 33*, 1327-1336.

Solow, A., & Gorelick, S. (1986). Estimating monthly streamflow values by cokriging. *Mathematical Geology, 18*(8), 785-809.

Spadavecchia, L. (2008). *Estimation of Landscape Carbon Budgets: Combining Geostatistical and Data Assimilation Approaches.* (Doctoral dissertation). The University of Edinburgh, Edinburgh.

Sporer, E., & Kangas, M. (2004). Shark Bay Prawn Managed Fishery Status Report. In J. W. Penn, W. J. Fletcher & F. Head (Eds.), *State of the Fisheries Report 2003/2004* (pp. 76-80): Department of Fisheries, Western Australia.

Sporer, E., & Kangas, M. (2005). Shark Bay Prawn Managed Fishery Status Report. In J. W. Penn, W. J. Fletcher & F. Head (Eds.), *State of the Fisheries Report 2004/2005* (pp. 86-91): Department of Fisheries, Western Australia.

Sporer, E., Kangas, M., & Brown, S. (2010). Shark Bay Prawn and Scallop Managed Fishery Status Report. In W. J. Fletcher & K. Santoro (Eds.), *State of the Fisheries Report 2009/2010* (pp. 107-116): Department of Fisheries, Western Australia.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging*. New York: Springer.

Stein, M. L. (2005). Space-time covariance functions. *Journal of the American Statistical Association, 100*(469), 310-321.

Telfer, C. (2010). *The Western Australian charter boat industry: working towards long-term sustainability.* (Master's thesis). Edith Cowan University, Perth.

Tjøstheim, D. (1978). A measure of association for spatial variables. *Biometrika, 65*(1), 109-114.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Cambridge: Addison-Wesley.

Walters, C. (2003). Folly and fantasy in the analysis of spatial catch rate data. *Canadian Journal of Fisheries and Aquatic Sciences, 60*, 1433-1436.

Yamamoto, J. K. (2007). On unbiased backtransform of lognormal kriging. *Computational Geosciences, 11*, 219-234.

APPENDICES

APPENDIX A: SUPPORTING PEER-REVIEWED CONFERENCE PUBLICATIONS

The following two peer-reviewed conference publications support the work in this thesis.

Denham, A., & Mueller, U. (2009). Space-time geostatistical analysis of King Prawn catch rate. In J. M. Ortiz & X. Emery (Eds.), GEOSTATS 2008 – Proceedings of the Eighth International Geostatistics Congress: Mining Engineering Department, University of Chile.

Denham, A., & Mueller, U. (2010). Incorporating survey data to improve space-time geostatistical analysis of King Prawn catch rate. In P. M. Atkinson & C. D. Lloyd (Eds.), geoENV VII – Geostatistics for Environmental Applications. Dordrecht: Springer.

APPENDIX B: PRAWN FISHERY CLOSURE LINES 2001-2004

Table B-1. Shark Bay prawn trawling closures, 2001-2004.

Season	Closure Line/Area	Opens	Closes	Re-opens	Re-closes
2001	Carnarvon-Peron Line	16/04			
	Extended Nursery Area	16/04	01/08		
	Extended Nursery Area below 25º18'S	16/04	20/04	15/05	01/08
	Extended Nursery Area below 25º20'24"S	15/05	01/08		
	Torbay Line	01/08			
	Denham Line	14/03	01/05	01/08	
	Tiger Prawn Spawning Area	14/03	19/06		
	Northern Tiger Prawn Spawning Area	14/03	24/07		
2002	Carnarvon-Peron Line	11/04			
	Extended Nursery Area	06/05	01/08		
	Torbay Line	01/08			
	Denham Line	06/03	01/05	01/08	
	Tiger Prawn Spawning Area	06/03	23/06		
	Northern Tiger Prawn Spawning Area	06/03			
2003	Carnarvon-Peron Line	24/04			
	Extended Nursery Area	21/05	01/08		
	Torbay Line	01/08			
	41 Minute Line	01/08			
	Denham Line	01/04	01/05	01/08	
	Tiger Prawn Spawning Area	06/03	21/05		
2004	Carnarvon-Peron Line	24/04			
	Carnarvon Peron below 25º13'S	01/05			
	Extended Nursery Area	24/04	05/08		
	ENA below	01/05			
	41 Minute	08/08			
	Denham	18/03	01/04	08/08	
	Tiger Prawn Spawning Area	16/03	26/05		

APPENDIX C: DATES OF MOON PHASES 2001-2004

Table C-1. Dates of moon phases for seasons 2001-2004 (Phases: L=Last Quarter, N=New Moon, Q=First Quarter, F=Full Moon).

Season 2001		Season 2002		Season 2003		Season 2004	
Date	Phase	Date	Phase	Date	Phase	Date	Phase
3/03	Q	6/03	L	3/03	N	7/03	F
10/03	F	14/03	N	11/03	Q	14/03	L
17/03	L	22/03	Q	18/03	F	21/03	N
25/03	N	29/03	F	25/03	L	29/03	Q
1/04	Q	4/04	L	2/04	N	5/04	F
8/04	F	13/04	N	10/04	Q	12/04	L
15/04	L	20/04	Q	17/04	F	19/04	N
23/04	N	27/04	F	23/04	L	28/04	Q
1/05	Q	4/05	L	1/05	N	5/05	F
7/05	F	12/05	N	9/05	Q	11/05	L
15/05	L	20/05	Q	16/05	F	19/05	N
23/05	N	26/05	F	23/05	L	27/05	Q
30/05	Q	3/06	L	31/05	N	3/06	F
6/06	F	11/06	N	8/06	Q	10/06	L
14/06	L	18/06	Q	14/06	F	18/06	N
21/06	N	25/06	F	21/06	L	26/06	Q
28/06	Q	3/07	L	30/06	N	2/07	F
5/07	F	10/07	N	7/07	Q	9/07	L
14/07	L	17/07	Q	14/07	F	17/07	N
21/07	N	24/07	F	21/07	L	25/07	Q
27/07	Q	1/08	L	29/07	N	1/08	F
4/08	F	9/08	N	5/08	Q	8/08	L
12/08	L	15/08	Q	12/08	F	16/08	N
19/08	N	23/08	F	20/08	L	23/08	Q
26/08	Q	31/08	L	28/08	N	30/08	F
3/09	F	7/09	N	3/09	Q	6/09	L
11/09	L	14/09	Q	10/09	F	14/09	N
17/09	N	21/09	F	19/09	L	21/09	Q
24/09	Q	30/09	L	26/09	N	28/09	F
2/10	F	6/10	N	3/10	Q	6/10	L
10/10	L	13/10	Q	10/10	F	14/10	N
17/10	N	21/10	F	18/10	L	21/10	Q
24/10	Q	29/10	L	25/10	N	28/10	F

APPENDIX D: LUNAR WEEKS OF 2001-2004 FISHING SEASONS

Table D-1. Lunar weeks for the fishing months of seasons 2001-2004. Start and End represent day/month of respective year. Phase indicates the name of the lunar week denoting the associated month and phase i.e. **MarF** = **F**ull moon week of **Mar**ch.

Season 2001			Season 2002			Season 2003			Season 2004		
Start	End	Phase	Start	End	Phase	Start	End	Phase	Start	End	Phase
7/03	13/03	MarF	3/03	10/03	MarL	28/02	7/03	MarN	11/03	17/03	MarL
14/03	21/03	MarL	11/03	18/03	MarN	8/03	14/03	MarQ	18/03	25/03	MarN
22/03	28/03	MarN	19/03	25/03	MarQ	15/03	21/03	MarF	26/03	1/04	MarQ
29/03	4/04	MarQ	26/03	1/04	MarF	22/03	29/03	MarL	2/04	8/04	AprF
5/04	11/04	AprF	1/04	7/04	AprL	30/03	6/04	AprN	9/04	15/04	AprL
12/04	19/04	AprL	10/04	16/04	AprN	7/04	13/04	AprQ	16/04	22/04	AprN
20/04	27/04	AprN	17/04	23/04	AprQ	14/04	20/04	AprF	25/04	1/05	AprQ
28/04	4/05	AprQ	24/04	30/04	AprF	20/04	27/04	AprL	2/05	8/05	MayF
5/05	11/05	MayF	1/05	8/05	MayL	28/04	5/05	MayN	9/05	15/05	MayL
12/05	19/05	MayL	10/05	16/05	MayN	7/05	12/05	MayQ	16/05	23/05	MayN
20/05	26/05	MayN	17/05	23/05	MayQ	13/05	19/05	MayF	24/05	30/05	MayQ
27/05	2/06	MayQ	23/05	30/05	MayF	20/05	27/05	MayL	31/05	6/06	JunF
3/06	10/06	JunF	31/05	7/06	JunL	28/05	4/06	JunN	7/06	14/06	JunL
11/06	17/06	JunL	8/06	14/06	JunN	5/06	11/06	JunQ	15/06	22/06	JunN
18/06	24/06	JunN	15/06	21/06	JunQ	11/06	17/06	JunF	23/06	29/06	JunQ
25/06	1/07	JunQ	22/06	29/06	JunF	18/06	24/06	JunL	30/06	5/07	JulF
2/07	9/07	JulF	30/06	6/07	JulL	27/06	3/07	JulN	6/07	13/07	JulL
10/07	17/07	JulL	7/07	14/07	JulN	4/07	11/07	JulQ	14/07	21/07	JulN
18/07	24/07	JulN	13/07	20/07	JulQ	10/07	17/07	JulF	22/07	28/07	JulQ
24/07	1/08	JulQ	21/07	28/07	JulF	18/07	25/07	JulL	29/07	4/08	AugF
2/08	8/08	AugF	29/07	1/08	AugL	26/07	1/08	AugN	5/08	12/08	AugL
9/08	15/08	AugL	2/08	12/08	AugN	2/08	8/08	AugQ	13/08	19/08	AugN
16/08	22/08	AugN	12/08	19/08	AugQ	9/08	16/08	AugF	20/08	26/08	AugQ
23/08	30/08	AugQ	20/08	27/08	AugF	17/08	24/08	AugL	27/08	2/09	SepF
31/08	7/09	SepF	28/08	3/09	SepL	25/08	31/08	SepN	3/09	10/09	SepL
8/09	14/09	SepL	4/09	10/09	SepN	31/08	6/09	SepQ	11/09	17/09	SepN
15/09	21/09	SepN	11/09	17/09	SepQ	7/09	13/09	SepF	18/09	25/09	SepQ
22/09	28/09	SepQ	19/09	25/09	SepF	17/09	23/09	SepL	26/09	2/10	OctF
29/09	6/10	OctF	28/09	3/10	OctL	24/09	29/09	OctN	3/10	10/10	OctL
7/10	13/10	OctL	3/10	9/10	OctN	30/09	6/10	OctQ	11/10	17/10	OctN
14/10	20/10	OctN	10/10	17/10	OctQ	7/10	14/10	OctF	18/10	24/10	OctQ
21/10	28/10	OctQ	18/10	25/10	OctF	15/10	21/10	OctL	25/10	1/11	NovF

APPENDIX E: ALIGNMENT OF LUNAR WEEKS OF 2001-2004 FISHING SEASONS

Table E-1. Week numbers for seasons 2001-2004 indicating the alignment of lunar weeks across seasons. This was achieved by the alignment of the four moon phases (Last Quarter, New Moon, First Quarter and New Moon) and the approximate alignment of each calendar month across years.

Week Number	2001	2002	2003	2004
1	MarL	MarL		MarL
2	MarN	MarN	MarN	MarN
3	MarQ	MarQ	MarQ	MarQ
4	AprF	MarF	MarF	AprF
5	AprL	AprL	MarL	AprL
6	AprN	AprN	AprN	AprN
7	AprQ	AprQ	AprQ	AprQ
8	MayF	AprF	AprF	MayF
9	MayL	MayL	AprL	MayL
10	MayN	MayN	MayN	MayN
11	MayQ	MayQ	MayQ	MayQ
12	JunF	MayF	MayF	JunF
13	JunL	JunL	MayL	JunL
14	JunN	JunN	JunN	JunN
15	JunQ	JunQ	JunQ	JunQ
16	JulF	JunF	JunF	JulF
17	JulL	JulL	JunL	JulL
18	JulN	JulN	JulN	JulN
19	JulQ	JulQ	JulQ	JulQ
20	AugF	JulF	JulF	AugF
21	AugL	AugL	JulL	AugL
22	AugN	AugN	AugN	AugN
23	AugQ	AugQ	AugQ	AugQ
24	SepF	AugF	AugF	SepF
25	SepL	SepL	AugL	SepL
26	SepN	SepN	SepN	SepN
27	SepQ	SepQ	SepQ	SepQ
28	OctF	SepF	SepF	OctF
29	OctL	OctL	SepL	OctL
30	OctN	OctN	OctN	OctN
31	OctQ	OctQ	OctQ	OctQ

APPENDIX F: DAILY KING PRAWN CPUE

F.1 Daily CPUE aggregated by vessel, 2001

Figure F.1-1. CPUEvd, March-June 2001

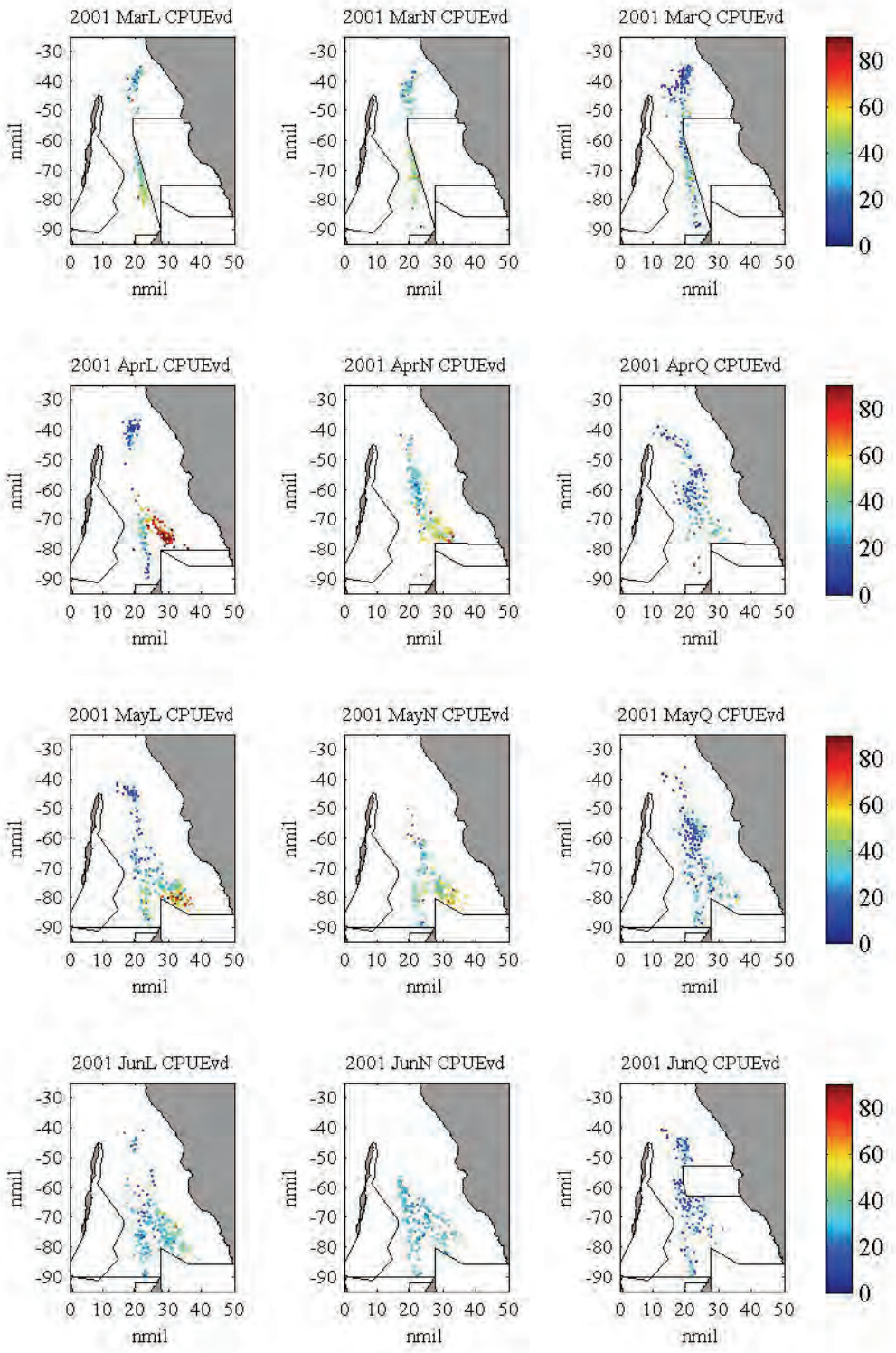
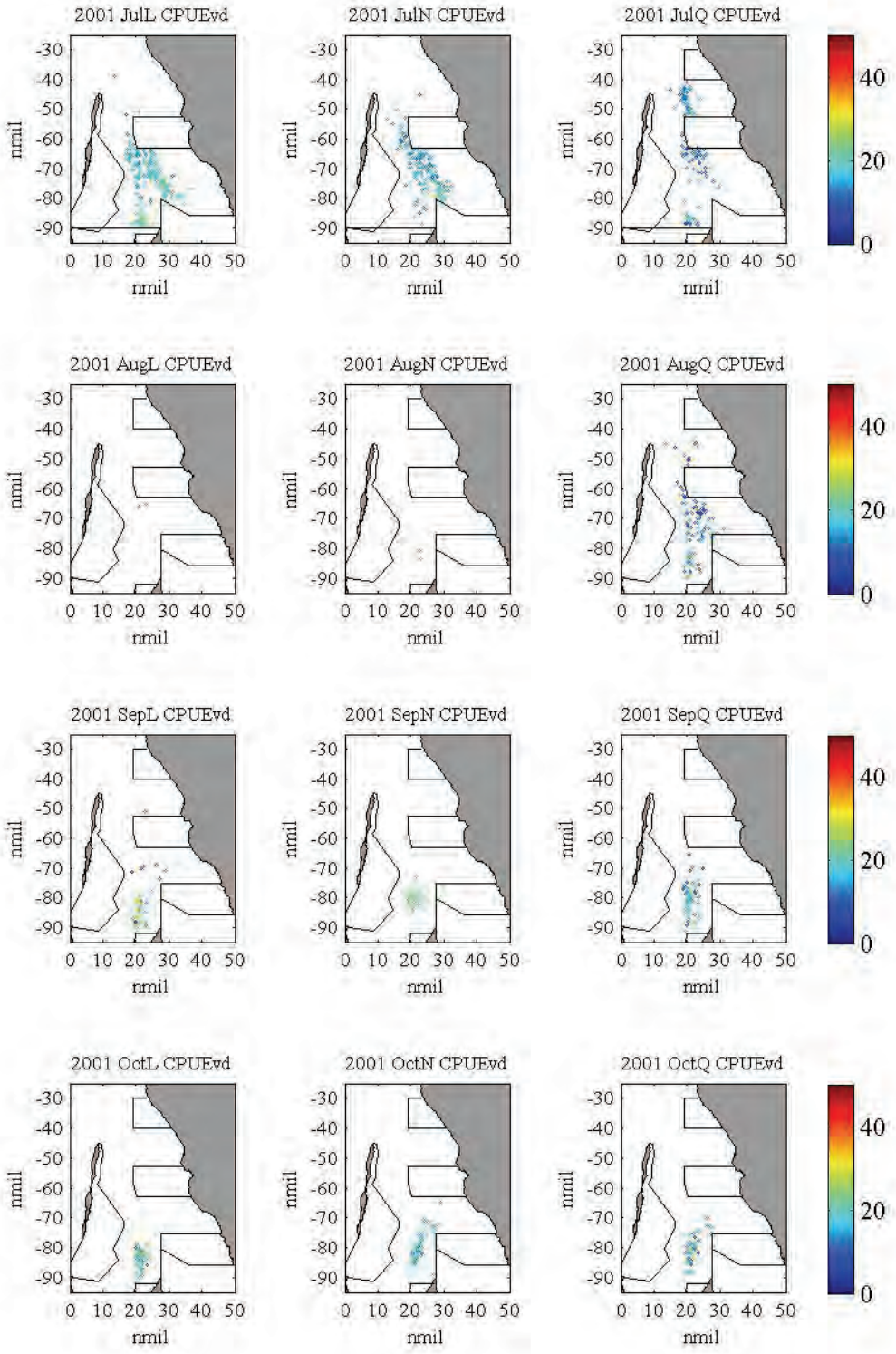


Figure F.1-2. CPUEvd, July-October 2001



F.2 Daily CPUE aggregated by vessel, 2002

Figure F.2-1. CPUEvd, March-June 2002

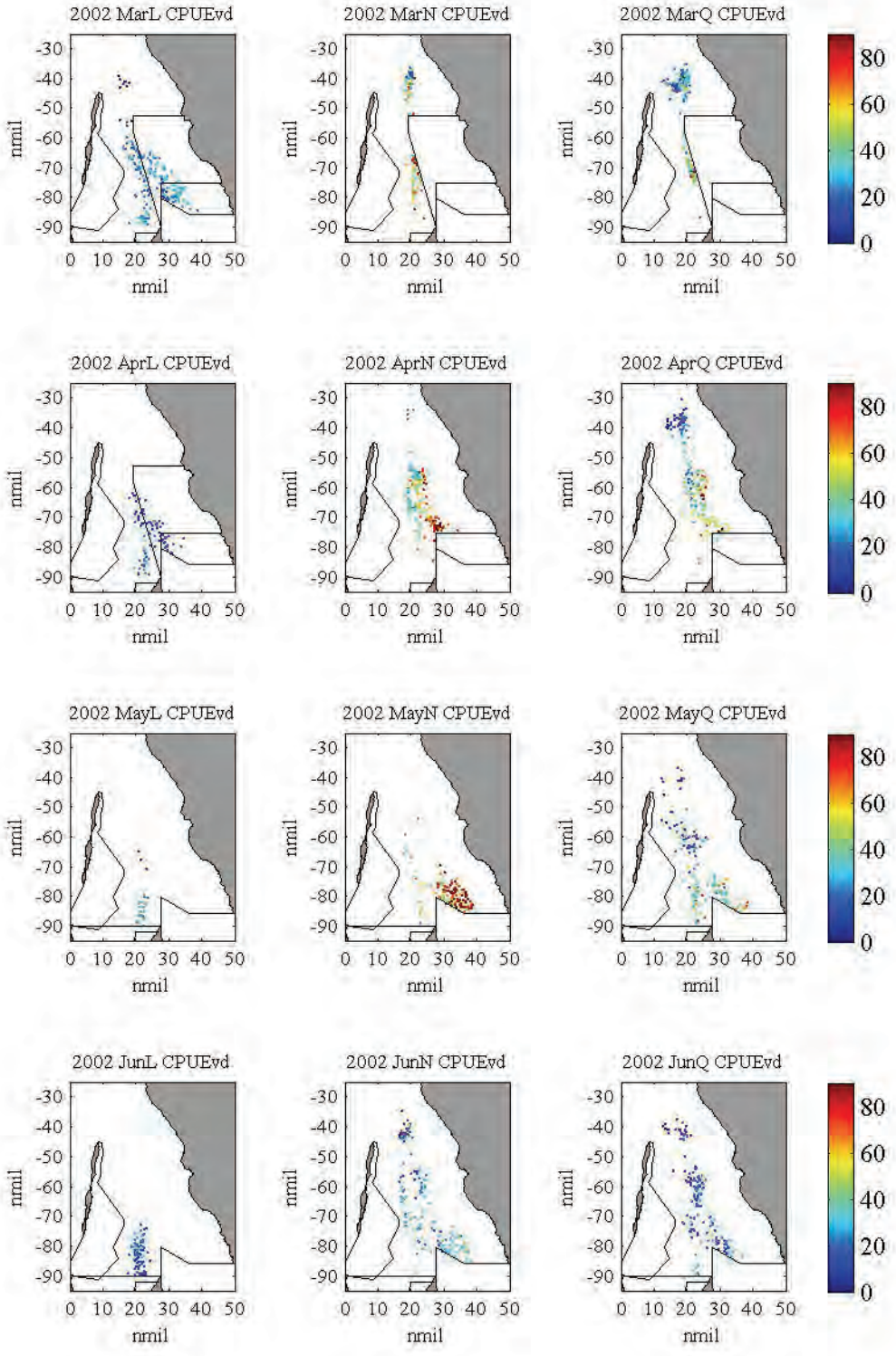
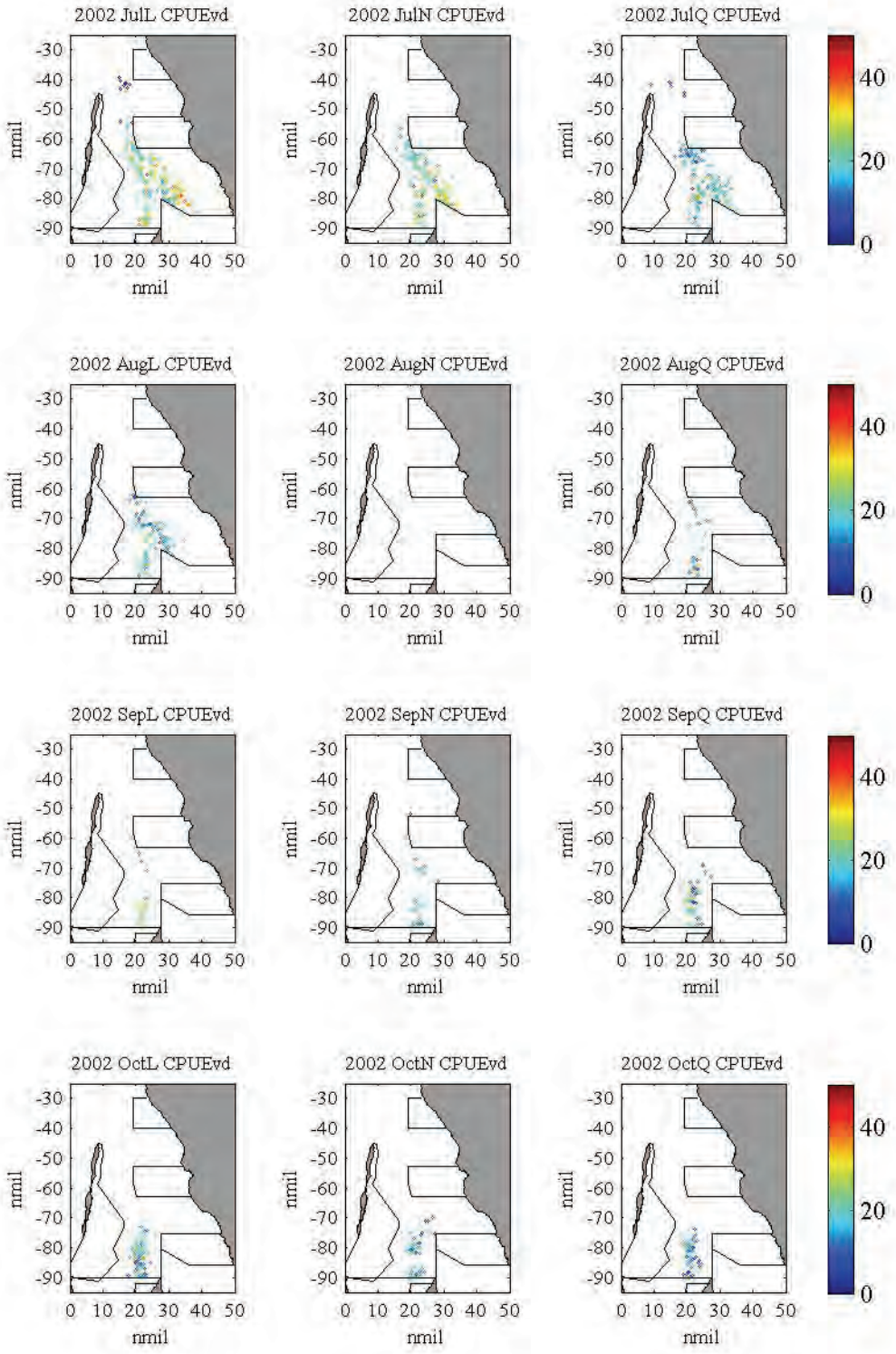


Figure F.2-2. CPUEvd, July-October 2002



F.3 Daily CPUE aggregated by vessel, 2003

Figure F.3-1. CPUEvd, March-June 2003

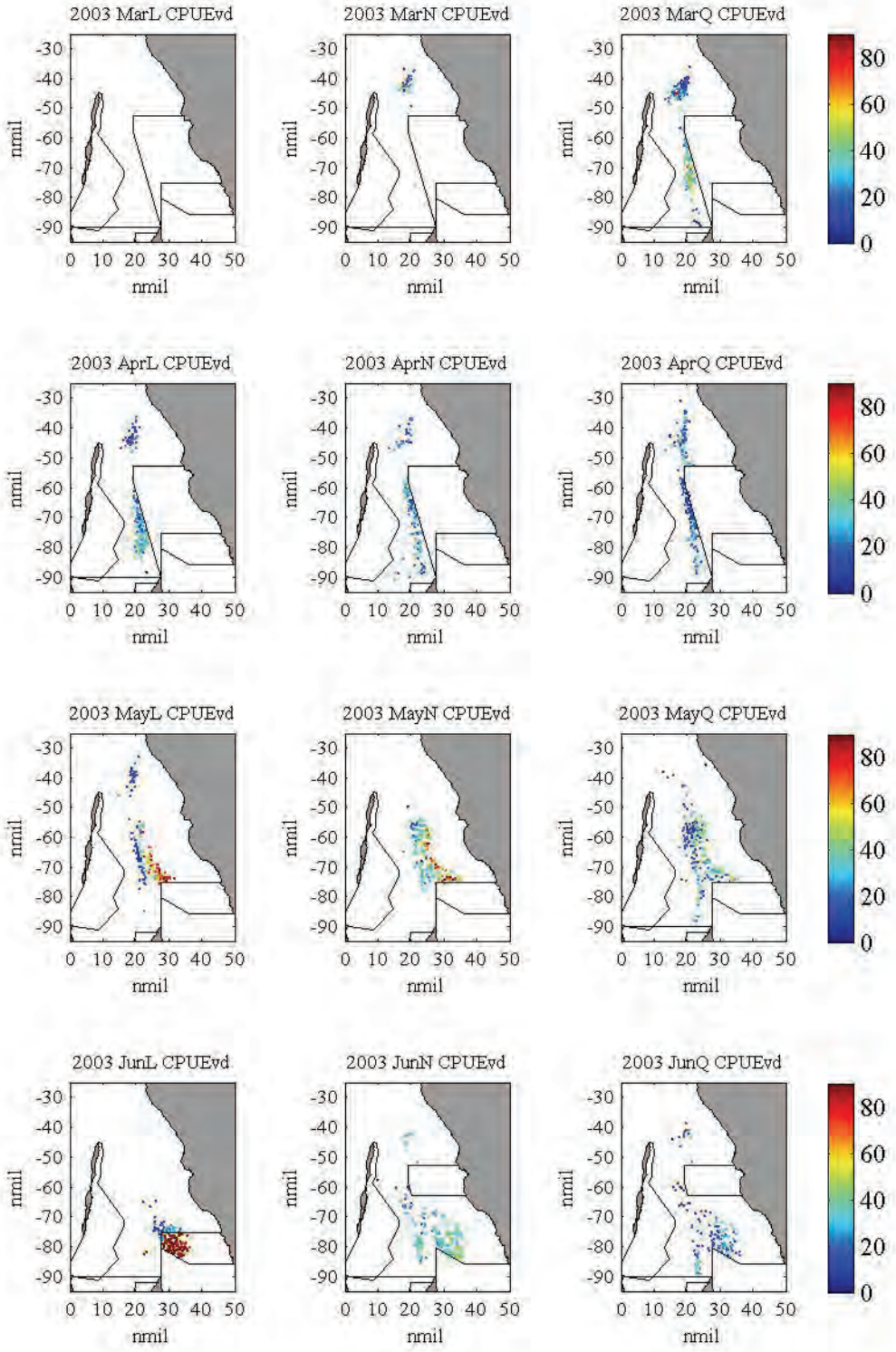
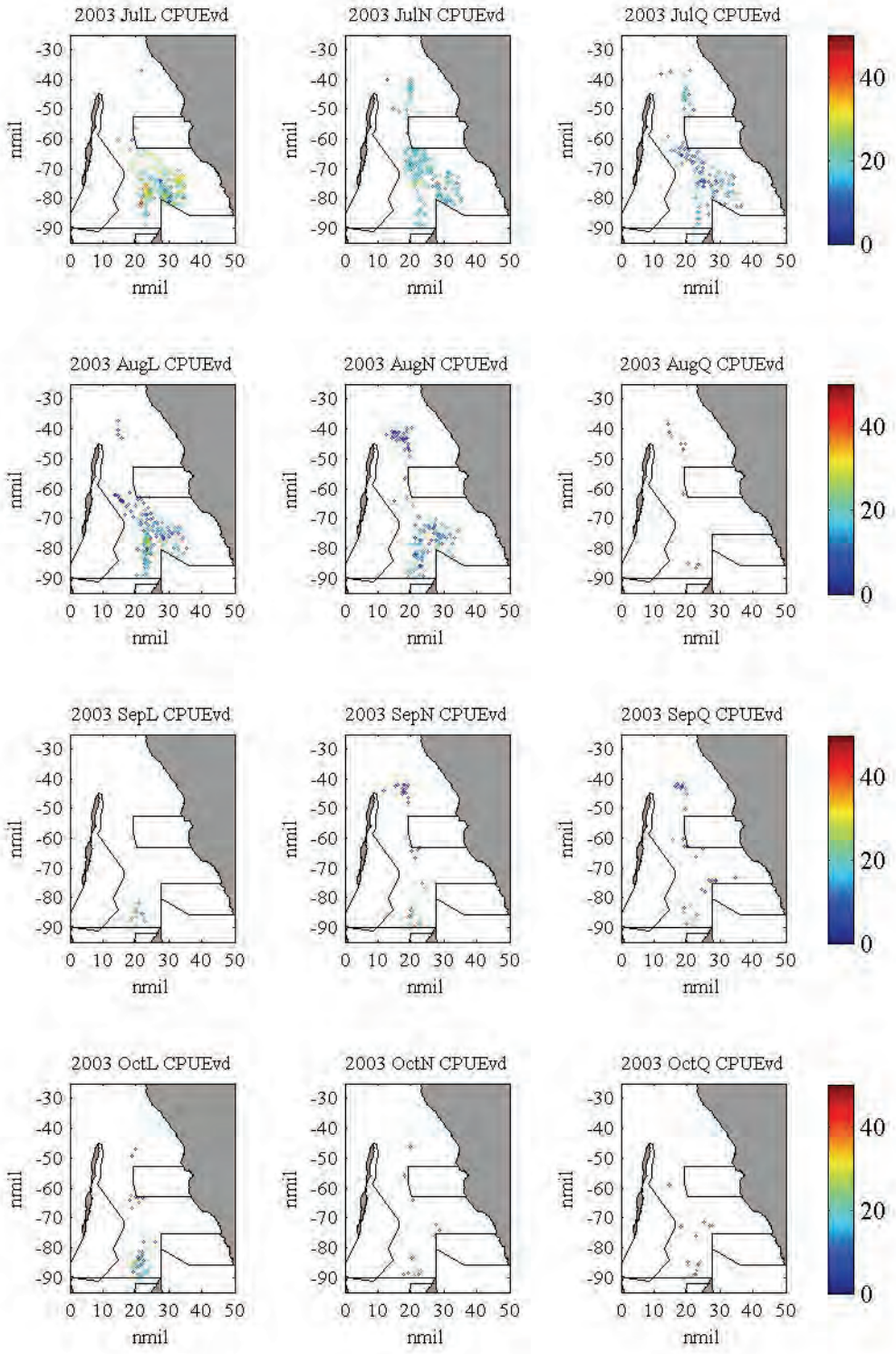


Figure F.3-2. CPUEvd, July-October 2003



F.4 Daily CPUE aggregated by vessel, 2004

Figure F.4-1. CPUEvd, March-June 2004

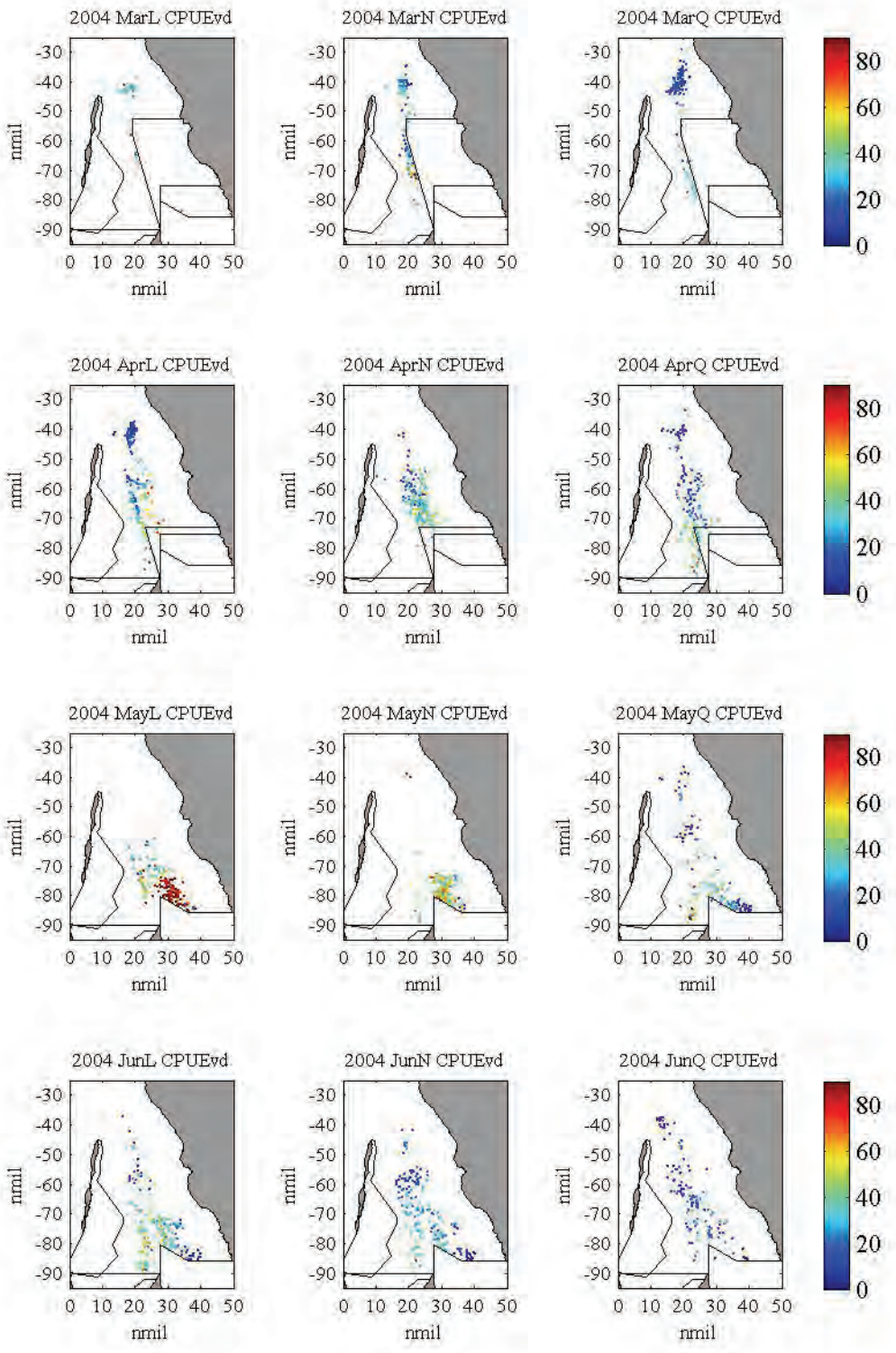
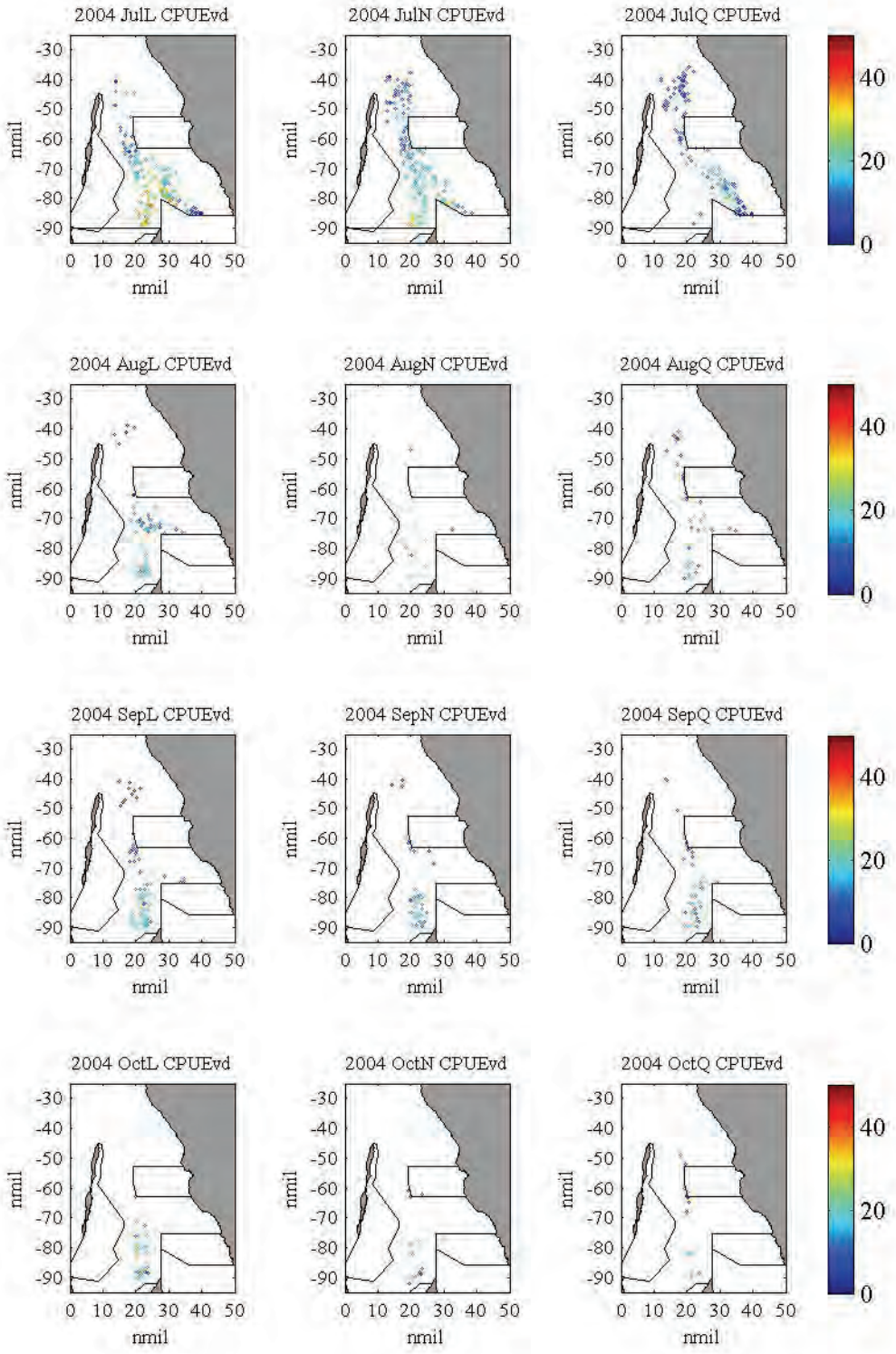


Figure F.4-2. CPUEvd, July-October 2004



F.5 Daily CPUE aggregated by grid, 2001

Figure F.5-1. CPUEgd, March-June 2001

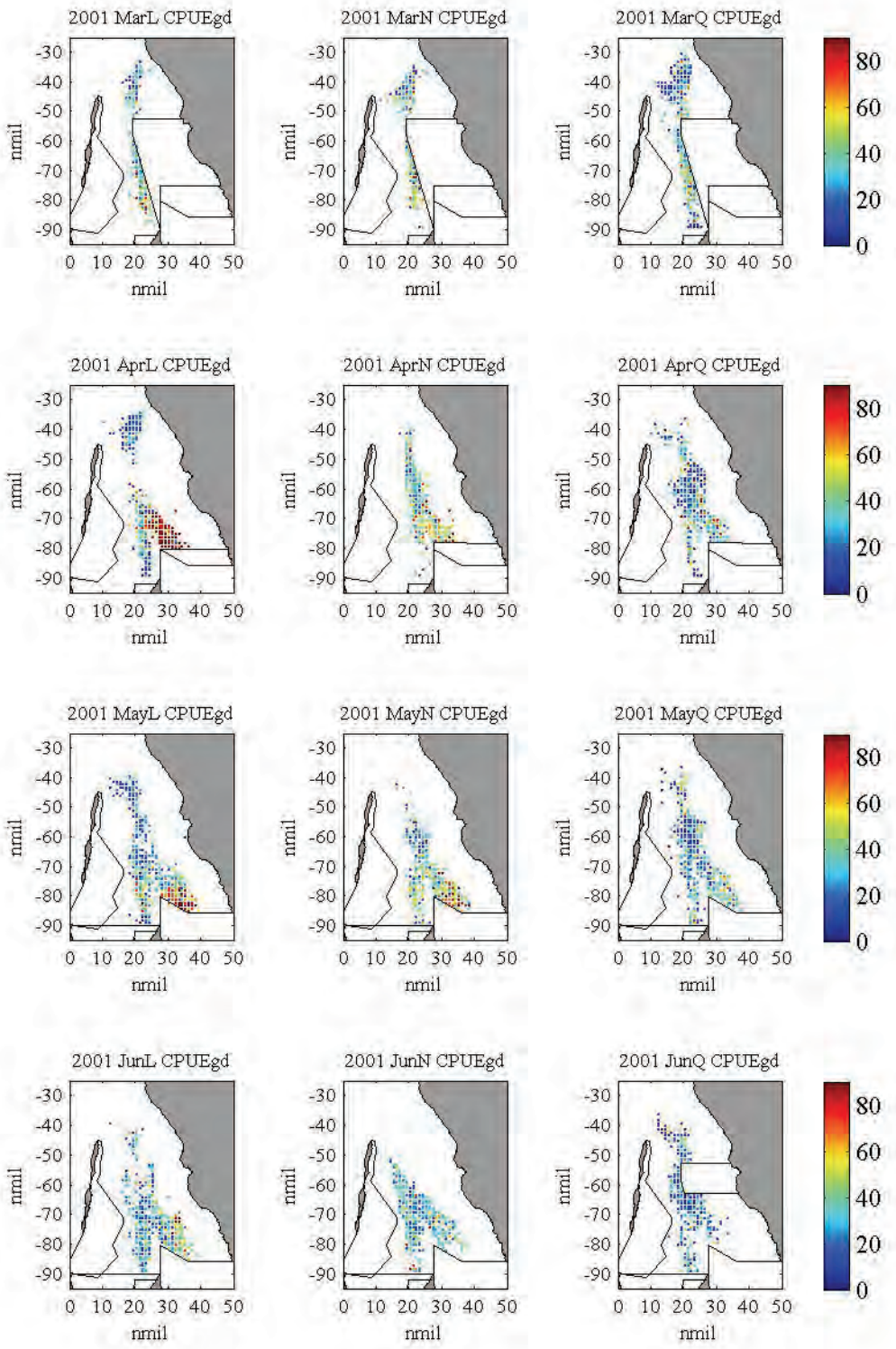
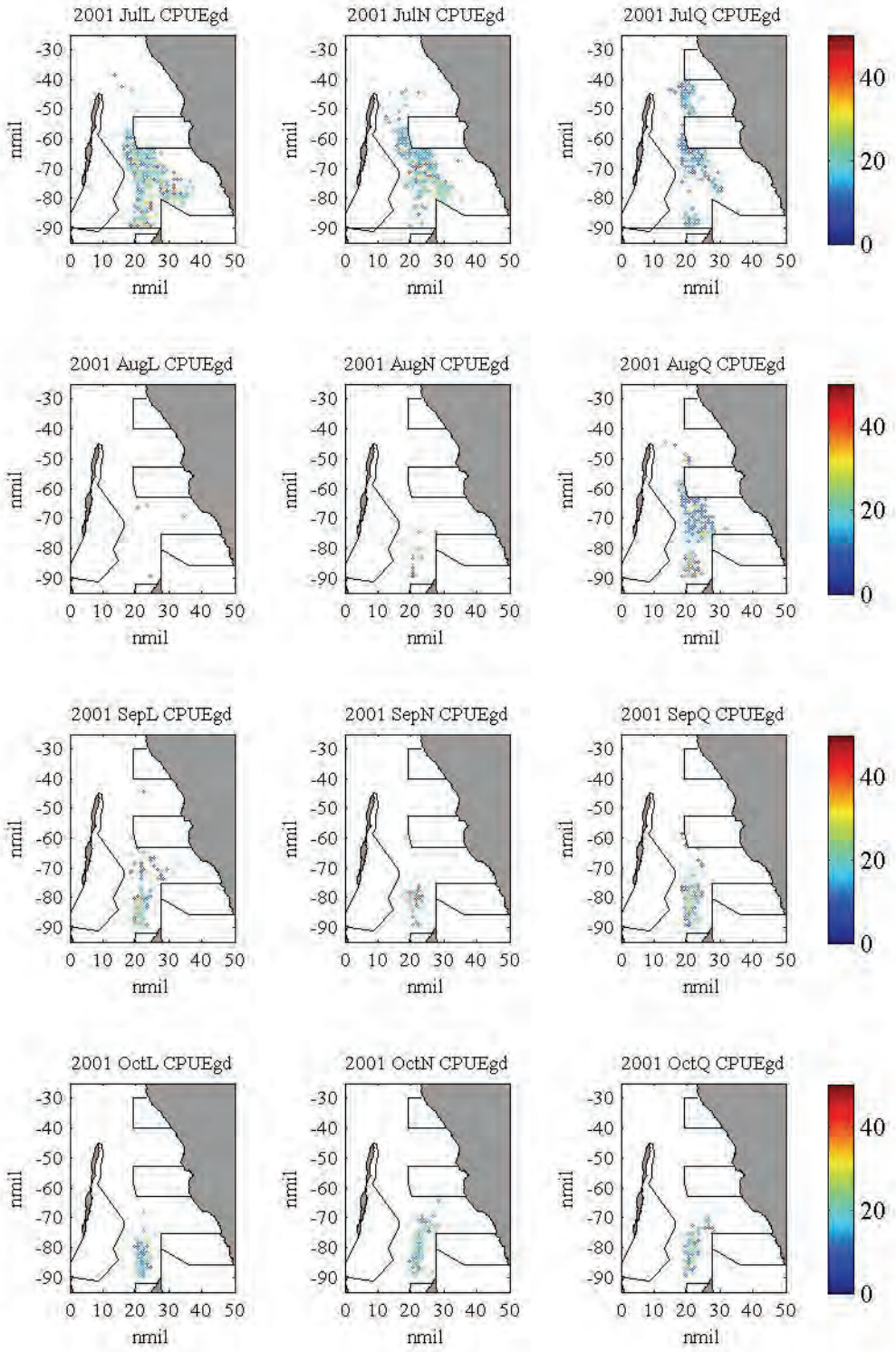


Figure F.5-2. CPUEvd, July-October 2001



F.6 Daily CPUE aggregated by grid, 2002

Figure F.6-1. CPUEgd, March-June 2002

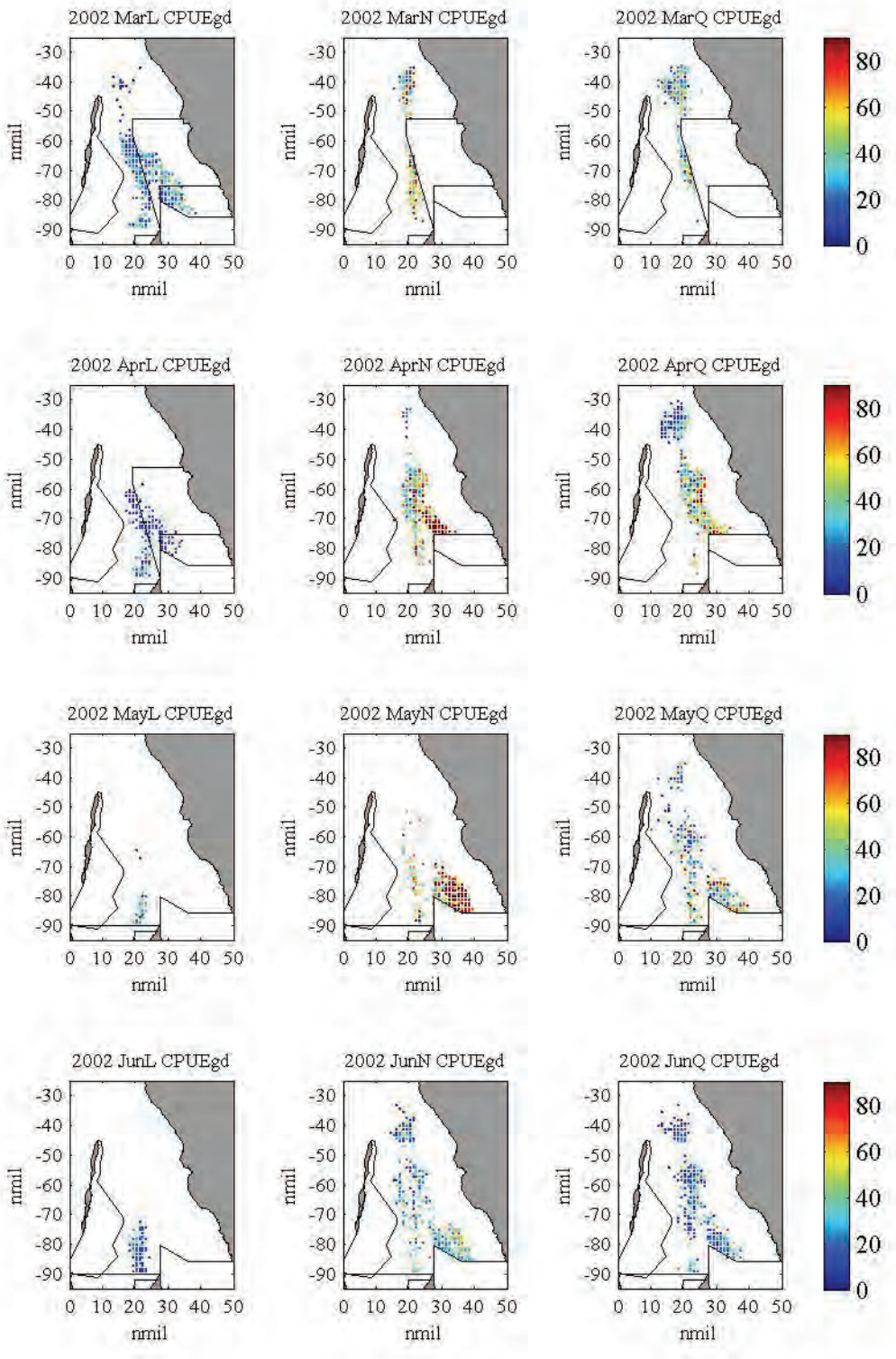
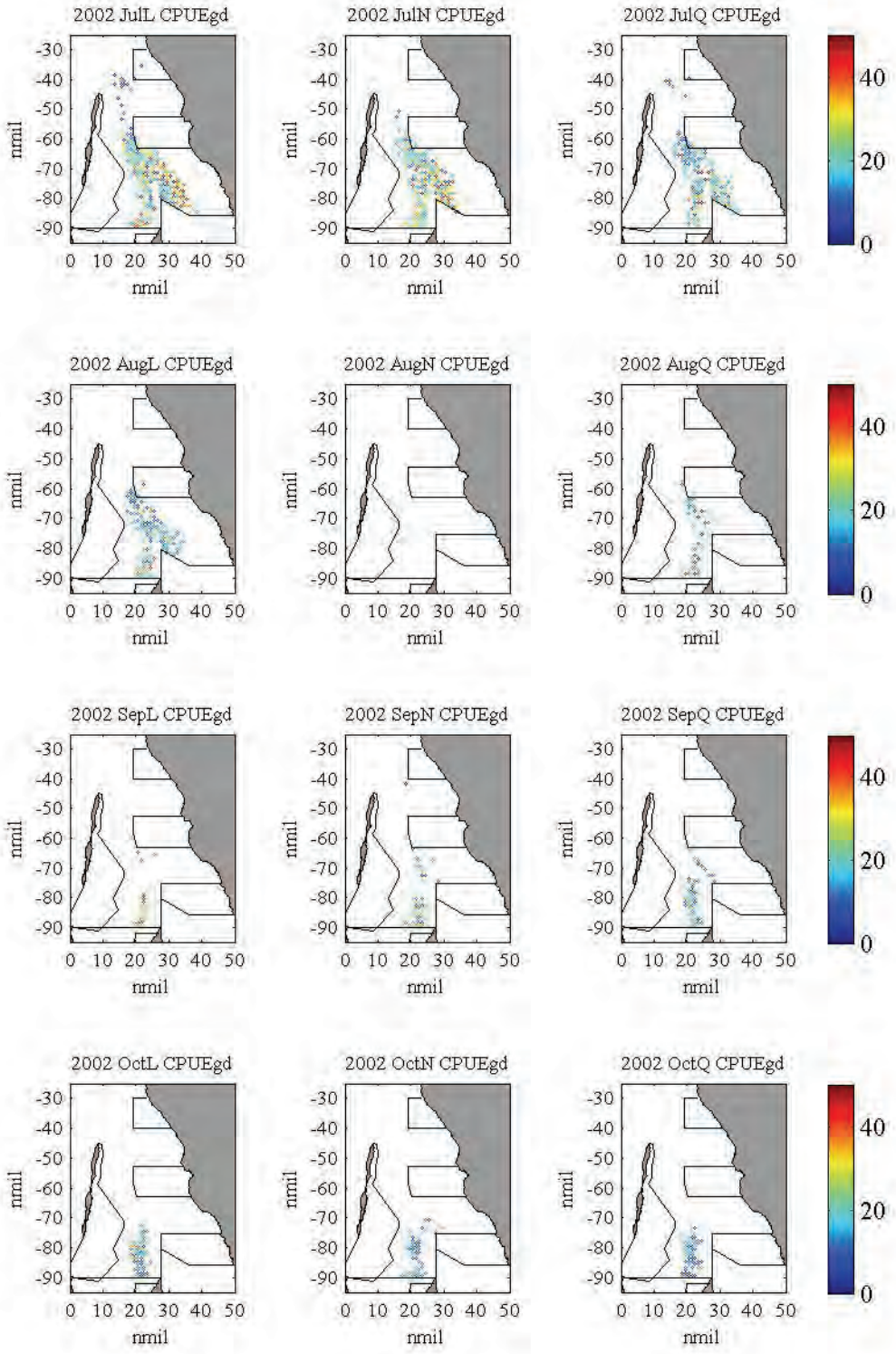


Figure F.6-2. CPUEgd, July-October 2002



F.7 Daily CPUE aggregated by grid, 2003

Figure F.7-1. CPUEgd, March-June 2003

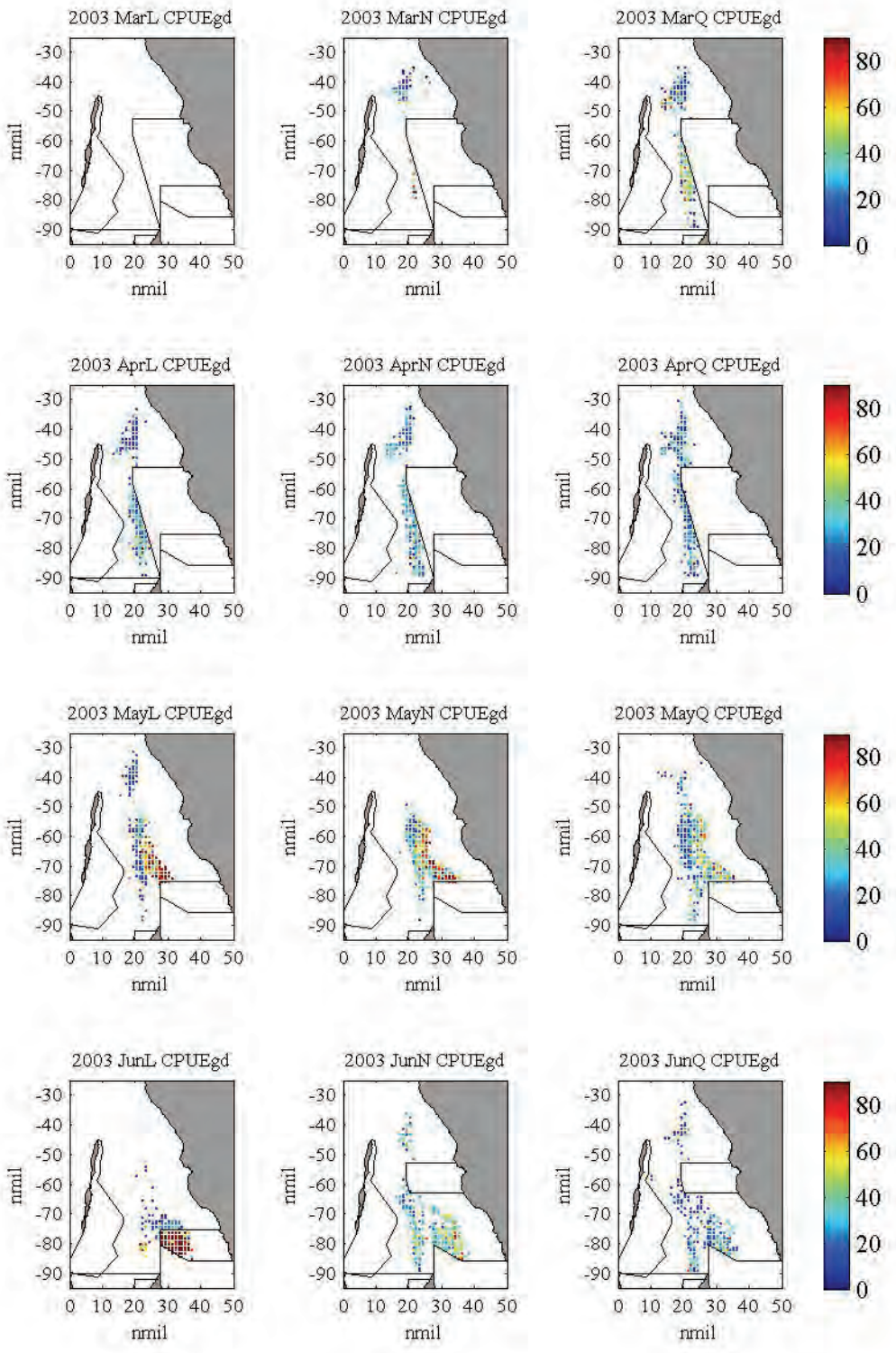
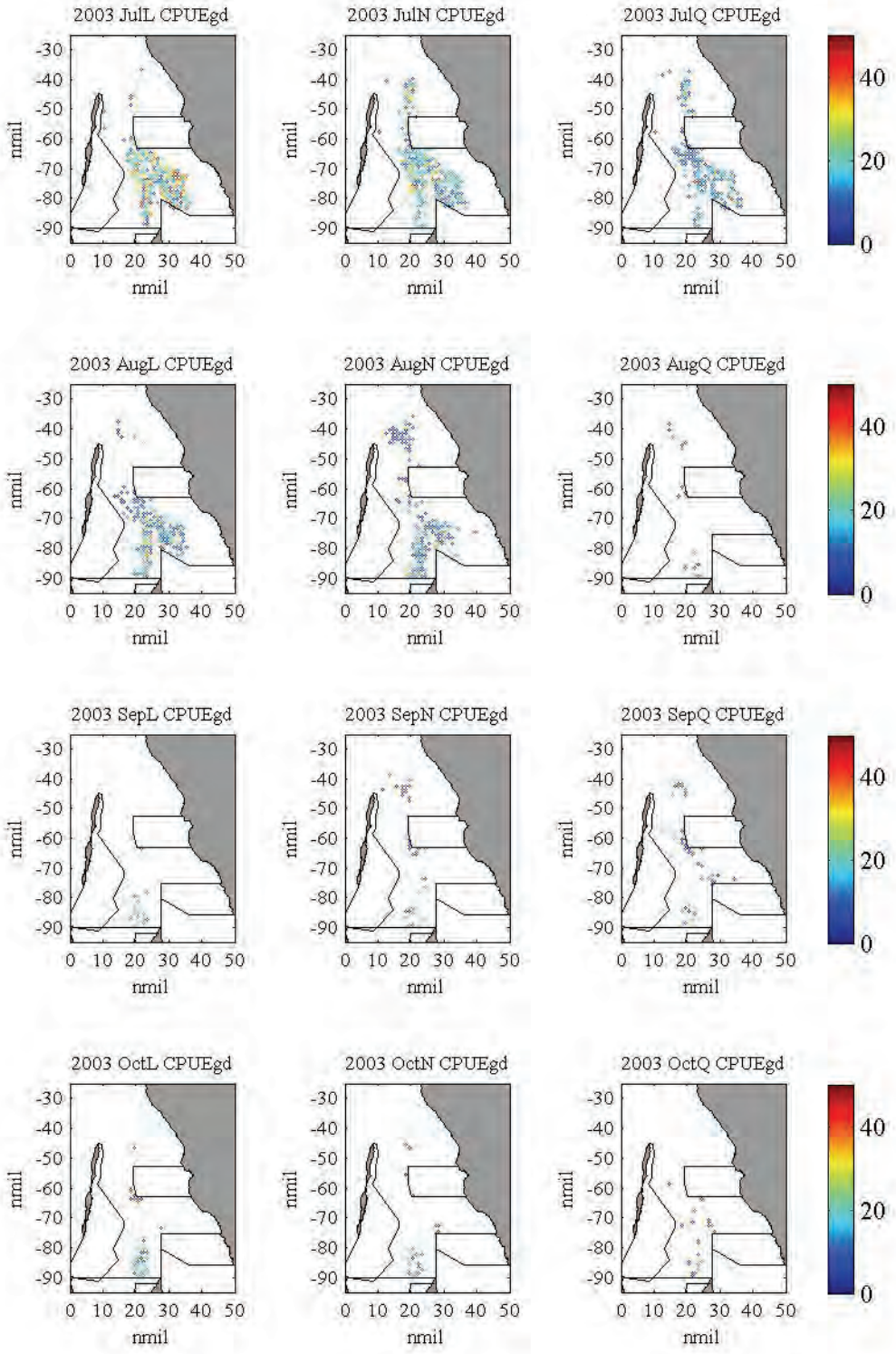


Figure F.7-2. CPUEgd, July-October 2003



F.8 Daily CPUE aggregated by grid, 2004

Figure F.8-1. CPUEgd, March-June 2004

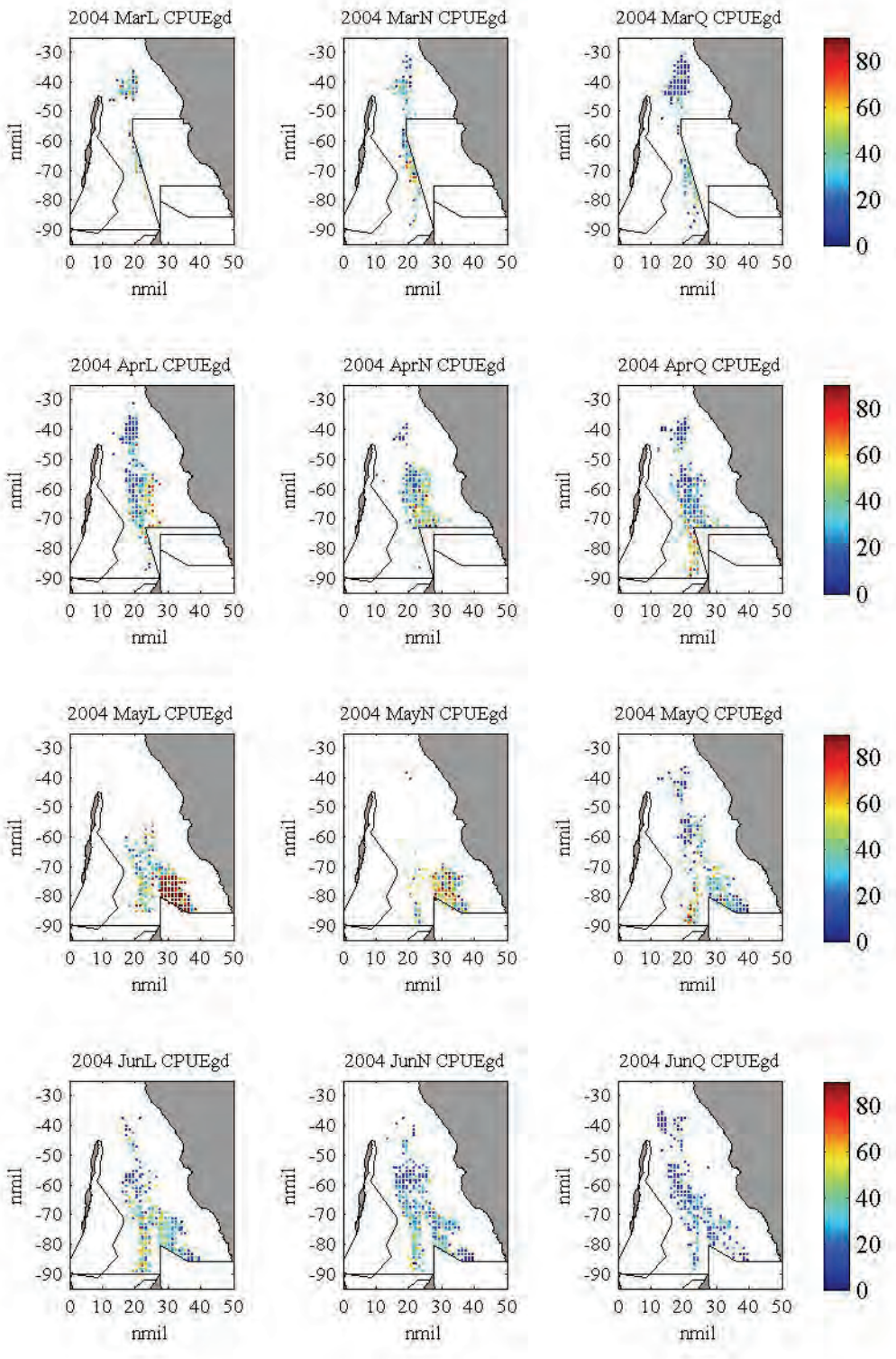
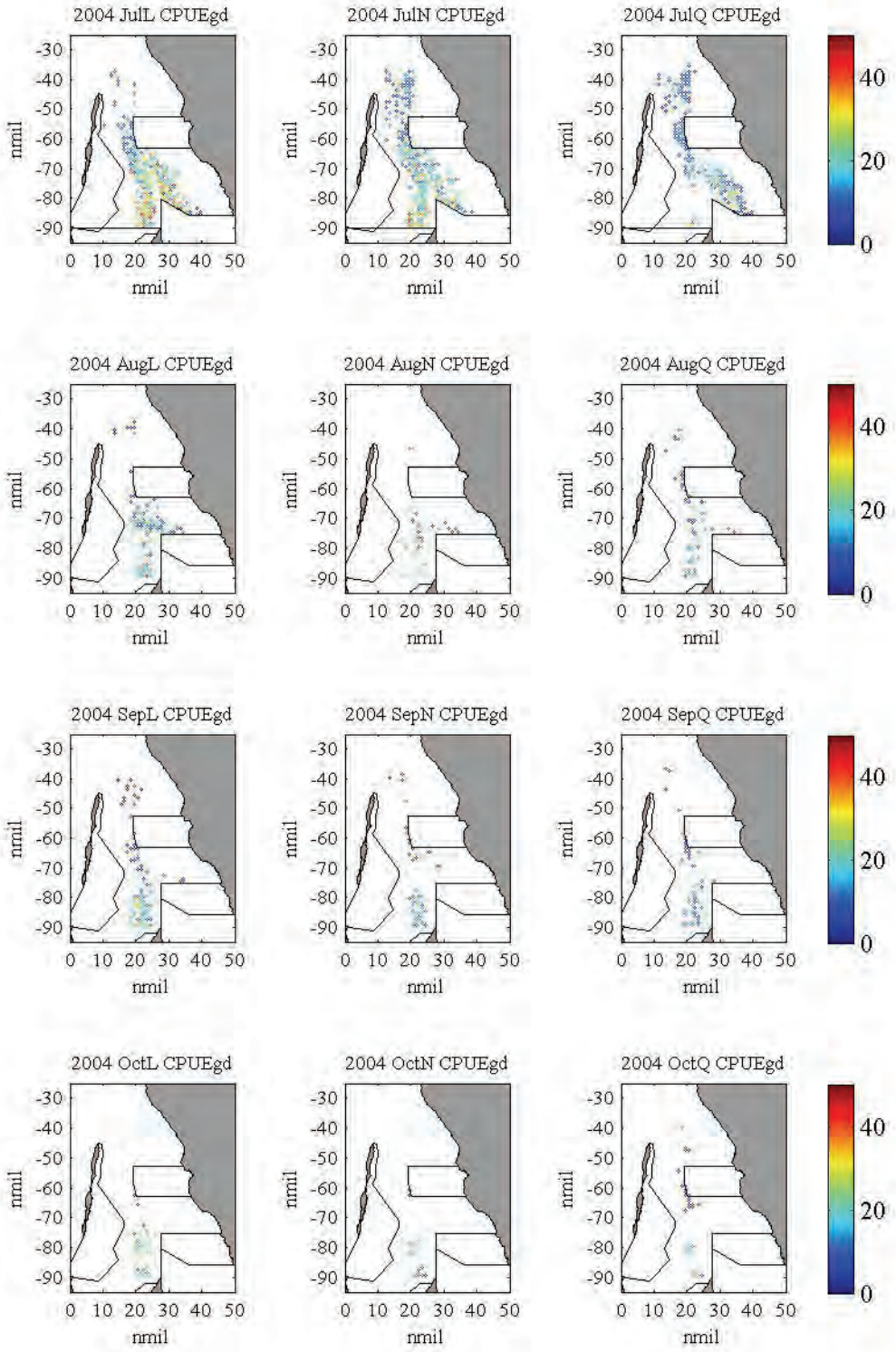


Figure F.8-2. CPUEgd, July-October 2004



APPENDIX G: PARAMETER FILES

G.1 Example parameter file for *gamv.exe*

```
                  Parameters for GAMV
                  **********************

START OF PARAMETERS:
DataGwW12001.dat                  \data file
2   3   0                         \column for x,y,z coordinates
3   11  15  16                    \nvar; column numbers...
-900.     1.0e21                  \tmin, tmax (trimming limits)
EVGwW1_2001.var                   \output file for variograms
40                                \nlag - number of spatial lags
1                                 \lag - separation distance
0.5                               \lag tolerance
1                                 \ndir - number of directions
0.0  90.0 50.0   0.0  90.0  50.0  \azm,atol,bandh,dip,dtol,bandv
0                                 \standardize sills?(0=no,1=yes)
3                                 \number of variograms
1   1   1                         \tail, head, variogram type
2   2   1                         \tail, head, variogram type
3   3   1                         \tail, head, variogram type
```

G.2 Example parameter file for *gamvmod.exe*

```
                  Parameters for GAMVMOD
                  **********************

START OF PARAMETERS:
Data2001.dat                      \data file
2   3   4                         \column for x,y, t coordinates
1   8                             \nvar; column numbers...
-9000.     1.0e21                 \tmin, tmax (trimming limits)
EV2001.var                        \output file for variograms
80                                \nlag - the number of spatial lags
0.5                               \xlag - unit separation distance
0.25                              \xltol- lag tolerance
60                                \ntlag - the number of temporal lags
1                                 \tlag - unit separation distance
3                                 \ndir - number of directions
   0.0  90.0 15000  0 0 0         \azm,atol,bandh,dip,dtol,bandv
   0.0  22.5 15000  0 0 0         \azm,atol,bandh,dip,dtol,bandv
  90.0  22.5 15000  0 0 0         \azm,atol,bandh,dip,dtol,bandv
  0                               \standardize sills? (0=no, 1=yes)
1                                 \number of variograms
1   1   1                         \tail, head, variogram type
```

G.3 Example parameter file for *kt3dnew.exe*

```
                  Parameters for KT3D
                  *******************

START OF PARAMETERS:
Data2004_Mar.dat                  \file with data
2   3   4   20   0                \   columns for X, Y, Z, var, sec var
-2000     1.0e21                  \   trimming limits
1                                 \option: 0=grid, 1=cross, 2=jackknife
Data2004_Mar.dat                  \file with jackknife data
2   3   4   20   0                 \   columns for X,Y,Z,vr and sec var
```

```
1                               \debugging level: 0,1,2,3
2004Marc.dbg                    \file for debugging output
2004Marc.dat                    \file for kriged output
31    10   1                    \nx,xmn,xsiz
61   -90   1                    \ny,ymn,ysiz
5   1096   1                    \nz,zmn,zsiz
1   1   1                       \x,y and z block discretization
4  20                           \min, max data for kriging
0                               \max per octant (0-> not used)
20   20   30                    \maximum search radii
0.0  0.0  0.0                   \   angles for search ellipsoid
1   0                           \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0               \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                               \0, variable; 1, estimate trend
nodata.dat                      \gridded file with drift/mean
0                               \  column number in gridded file
3  0.07                         \nst, spatial nugget effect
1  0.03  0.0   0.0   0.0        \it,cc,ang1,ang2,ang3
      1    1    1               \a_hmax, a_hmin, a_vert
1  0.05  0.0   0.0   0.0        \it,cc,ang1,ang2,ang3
      20   20   20              \a_hmax, a_hmin, a_vert
7  0.15  0.0   0.0   0.0        \it,cc,ang1,ang2,ang3
      12   12   12              \a_hmax, a_hmin, a_vert
0.00   0.299                    \temporal nugget, global sill
1                               \0 = product, 1 =sum-product
```

G.4 Example parameter file for *sgsimtemp.exe*

```
        Parameters for SGSIMTEMP
        ************************

START OF PARAMETERS:
Data2004_MarJunn.dat            \file with data
2  3  4  15  0  0               \  columns for X,Y,Z,vr,wt,sec.var
-1.0     1.0e21                 \  trimming limits
1                               \transform the data (0=no, 1=yes)
sgsimMarJun.trn                 \  file for output trans table
0                               \  consider ref. dist (0=no, 1=yes)
nodata.out                      \  file with ref. dist distribution
1  2                            \  columns for vr and wt
0.0    2.0                      \  zmin,zmax(tail extrapolation)
1      0.0                      \  lower tail option, parameter
4      5.0                      \  upper tail option, parameter
1                               \debugging level: 0,1,2,3
simJul.dbg                      \file for debugging output
simJul.dat                      \file for simulation output
100                             \number of realizations to generate
31    10   1                    \nx,xmn,xsiz
61   -90   1                    \ny,ymn,ysiz
8   1219   1                    \nz,zmn,zsiz
69069                           \random number seed
4    20                         \min and max original data for sim
6                               \number of simulated nodes to use
0                               \assign data to nodes (0=no, 1=yes)
0    3                          \multiple grid search (0=no,1=yes),num
0                               \maximum data per octant (0=not used)
20   20   30                    \maximum search radii (hmax,hmin,vert)
0.0  0.0  0.0                   \   angles for search ellipsoid
0    0.0  1.0                   \ktype: 0=SK,1=OK,2=LVM,3=EXDR,4=COLC
nodata.dat                      \file with LVM,EXDR,or COLC variable
1                               \column for secondary variable
```

```
3   0.20                        \nst, spatial nugget effect
1   0.04  0.0   0.0   0.0       \it,cc,ang1,ang2,ang3
        3     3     3           \a_hmax, a_hmin, a_vert
1   0.30  0.0   0.0   0.0       \it,cc,ang1,ang2,ang3
        15    15    15          \a_hmax, a_hmin, a_vert
7   0.52  0.0   0.0   0.0       \it,cc,ang1,ang2,ang3
        10    10    10          \a_hmax, a_hmin, a_vert
0.00   1.00                     \temporal nugget, global sill
1                               \0 = product, 1 =sum-product
```

APPENDIX H: SPATIAL VARIOGRAPHY

H.1 Spatial variography of weekly data sets of *CPUEsvd* and *CPUEsgd*

Figure H.1-1. Experimental semivariograms of *CPUEsvd* and *CPUEsgd* for Season 2001

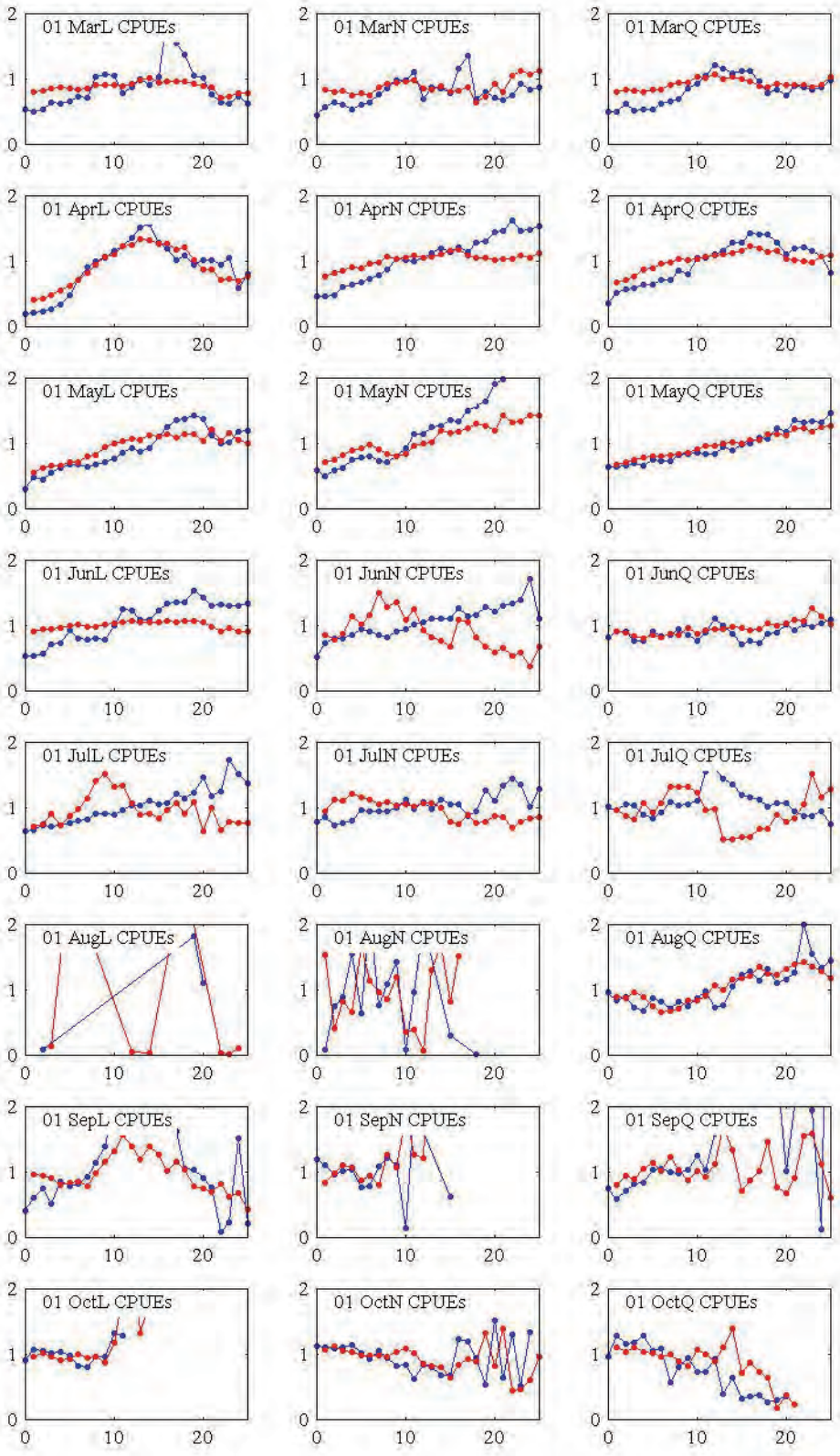


Figure H.1-2. Experimental semivariograms of *CPUEsvd* and *CPUEsgd* for Season 2002

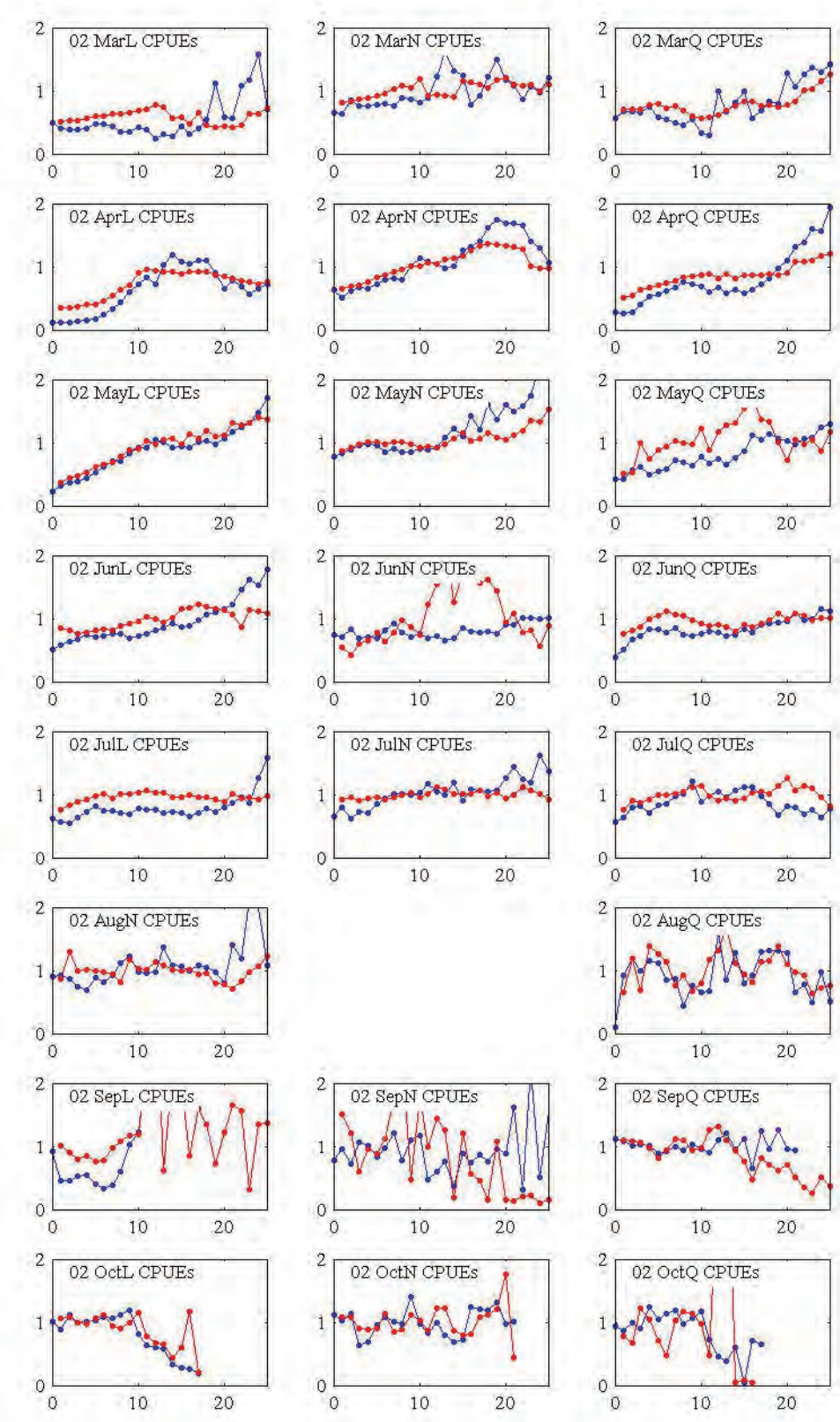
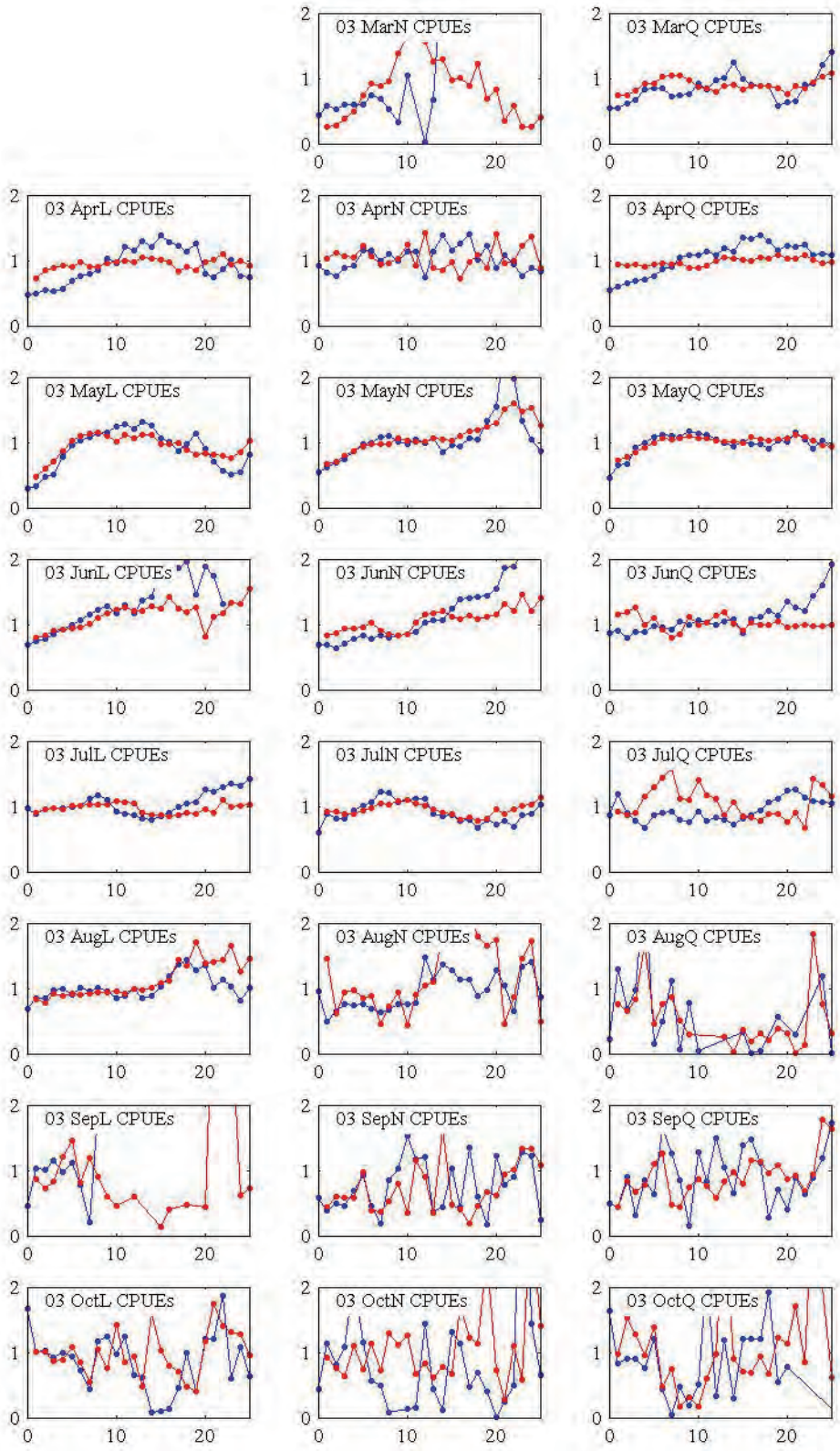


Figure H.1-3. Experimental semivariograms of *CPUEsvd* and *CPUEsgd* for Season 2003



H.2 Spatial semivariogram models of *CPUEs* variables

Table H.2-1. Spatial semivariogram model parameters of weekly *CPUEsvd* data sets.

	2001			2002			2003			2004		
Wk	Nugget	Range	Sill	Nugget	Range	Sill	Nugget	Range	Sill	Nugget	Range	Sill
1	0.07	16	0.08	0.08	8	0.07	0.00	10	0.08	0.10	15	0.09
2	0.04	15	0.06	0.09	5	0.07	0.15	15	0.25	0.10	9	0.23
3	0.08	15	0.14	0.12	5	0.07	0.30	15	0.34	0.03	20	0.14
5	0.10	15	0.60	0.01	5 20	0.01 0.10	0.03	14	0.04	0.01	8 20	0.07 0.10
6	0.05	12	0.10	0.07	10	0.14	0.03	7	0.02	0.08	11	0.05
7	0.07	15	0.11	0.07	10	0.15	0.05	11	0.05	0.08	22	0.23
9	0.09	18	0.23	0.20	10	0.50	0.10	8	0.34	0.15	10	0.41
10	0.04	12	0.08	0.19	3 20	0.03 0.1	0.19	8	0.16	0.12	9	0.05
11	0.11	20	0.09	0.10	5 25	0.02 0.20	0.11	5	0.16	0.12	16	0.35
13	0.04	13	0.04	0.03	2 20	0.03 0.04	0.80	6	0.43	0.05	16	0.12
14	0.02	2/15	0.01 0.01	0.04	5 20	0.01 0.01	0.06	15	0.04	0.03	15	0.13
15	0.11	25	0.03	0.04	5	0.04	0.12	10	0.03	0.04	2 18	0.03 0.09
17	0.02	15	0.01	0.03	7	0.03	0.05	5	0.01	0.04	20	0.08
18	0.02	10	0.01	0.03	10	0.04	0.03	5	0.02	0.02	28	0.09
19	0.09	10	0.00	0.07	10	0.05	0.07	20	0.04	0.04	5 20	0.09 0.07
21	0.32	10	0.00	0.06	10	0.00	0.09	8	0.04	0.04	20	0.08
22	0.02	5	0.11	2.00	5	0.00	0.07	15	0.09	0.31	1	0.00
23	0.15	15	0.06	0.02	1 15	0.20 0.10	0.09	2	0.28	0.05	2 20	0.25 0.07
25	0.10	8	0.20	0.06	10	0.00	0.10	1	0.24	0.08	25	0.22
26	0.11	10	0.00	0.06	4	0.03	0.40	20	0.80	0.10	20	0.12
27	0.09	6	0.11	0.34	6	0.00	0.10	10	0.32	0.10	2	0.33
29	0.10	1	0.00	0.14	2 10	0.04 0.01	0.29	1	0.00	0.07	5	0.10
30	0.06	10	0.00	0.09	10	0.00	0.10	2	0.14	0.06	20	0.06
31	0.12	2	0.00	0.07	2	0.05	0.27	2	0.00	0.10	15	0.15

Table H.2-2. Spatial semivariogram model parameters of weekly *CPUEsgd* data sets.

	2001			2002			2003			2004		
Wk	Nugget	Range	Sill	Nugget	Range	Sill	Nugget	Range	Sill	Nugget	Range	Sill
1	0.07	16	0.08	0.08	8	0.07	0.00	10	0.08	0.10	15	0.09
2	0.04	15	0.06	0.09	5	0.07	0.15	15	0.25	0.10	9	0.23
3	0.08	15	0.14	0.12	5	0.07	0.30	15	0.34	0.03	20	0.14
5	0.10	15	0.60	0.01	5 20	0.01 0.10	0.03	14	0.04	0.01	8 20	0.07 0.10
6	0.05	12	0.10	0.07	10	0.14	0.03	7	0.02	0.08	11	0.05
7	0.07	15	0.11	0.07	10	0.15	0.05	11	0.05	0.08	22	0.23
9	0.09	18	0.23	0.20	10	0.50	0.10	8	0.34	0.15	10	0.41
10	0.04	12	0.08	0.19	3 20	0.03 0.1	0.19	8	0.16	0.12	9	0.05
11	0.11	20	0.09	0.10	5 25	0.02 0.20	0.11	5	0.16	0.12	16	0.35
13	0.04	13	0.04	0.03	2 20	0.03 0.04	0.80	6	0.43	0.05	16	0.12
14	0.02	2/15	0.01 0.01	0.04	5 20	0.01 0.01	0.06	15	0.04	0.03	15	0.13
15	0.11	25	0.03	0.04	5	0.04	0.12	10	0.03	0.04	2 18	0.03 0.09
17	0.02	15	0.01	0.03	7	0.03	0.05	5	0.01	0.04	20	0.08
18	0.02	10	0.01	0.03	10	0.04	0.03	5	0.02	0.02	28	0.09
19	0.09	10	0.00	0.07	10	0.05	0.07	20	0.04	0.04	5 20	0.09 0.07
21	0.32	10	0.00	0.06	10	0.00	0.09	8	0.04	0.04	20	0.08
22	0.02	5	0.11	2.00	5	0.00	0.07	15	0.09	0.31	1	0.00
23	0.15	15	0.06	0.02	1 15	0.20 0.10	0.09	2	0.28	0.05	2 20	0.25 0.07
25	0.10	8	0.20	0.06	10	0.00	0.10	1	0.24	0.08	25	0.22
26	0.11	10	0.00	0.06	4	0.03	0.40	20	0.80	0.10	20	0.12
27	0.09	6	0.11	0.34	6	0.00	0.10	10	0.32	0.10	2	0.33
29	0.10	1	0.00	0.14	2 10	0.04 0.01	0.29	1	0.00	0.07	5	0.10
30	0.06	10	0.00	0.09	10	0.00	0.10	2	0.14	0.06	20	0.06
31	0.12	2	0.00	0.07	2	0.05	0.27	2	0.00	0.10	15	0.15

H.3 Spatial variography of weekly data sets of *CPUEsNvd* and *CPUEsNgd*

Figure H.3-1. Experimental semivariograms of *CPUEsNvd* and *CPUEsNgd* for Season 2001

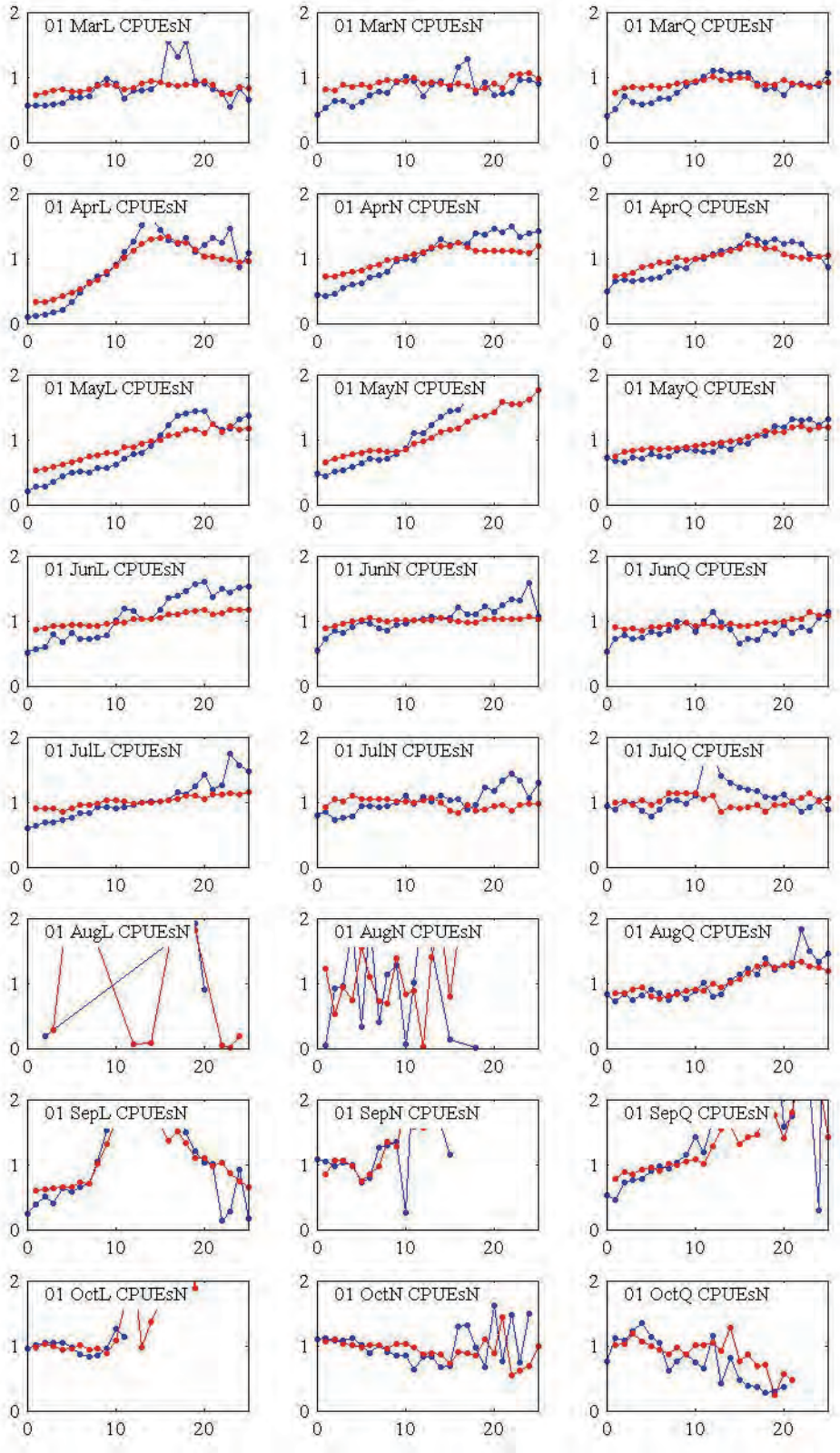


Figure H.3-2. Experimental semivariograms of *CPUEsNvd* and *CPUEsNgd* for Season 2002

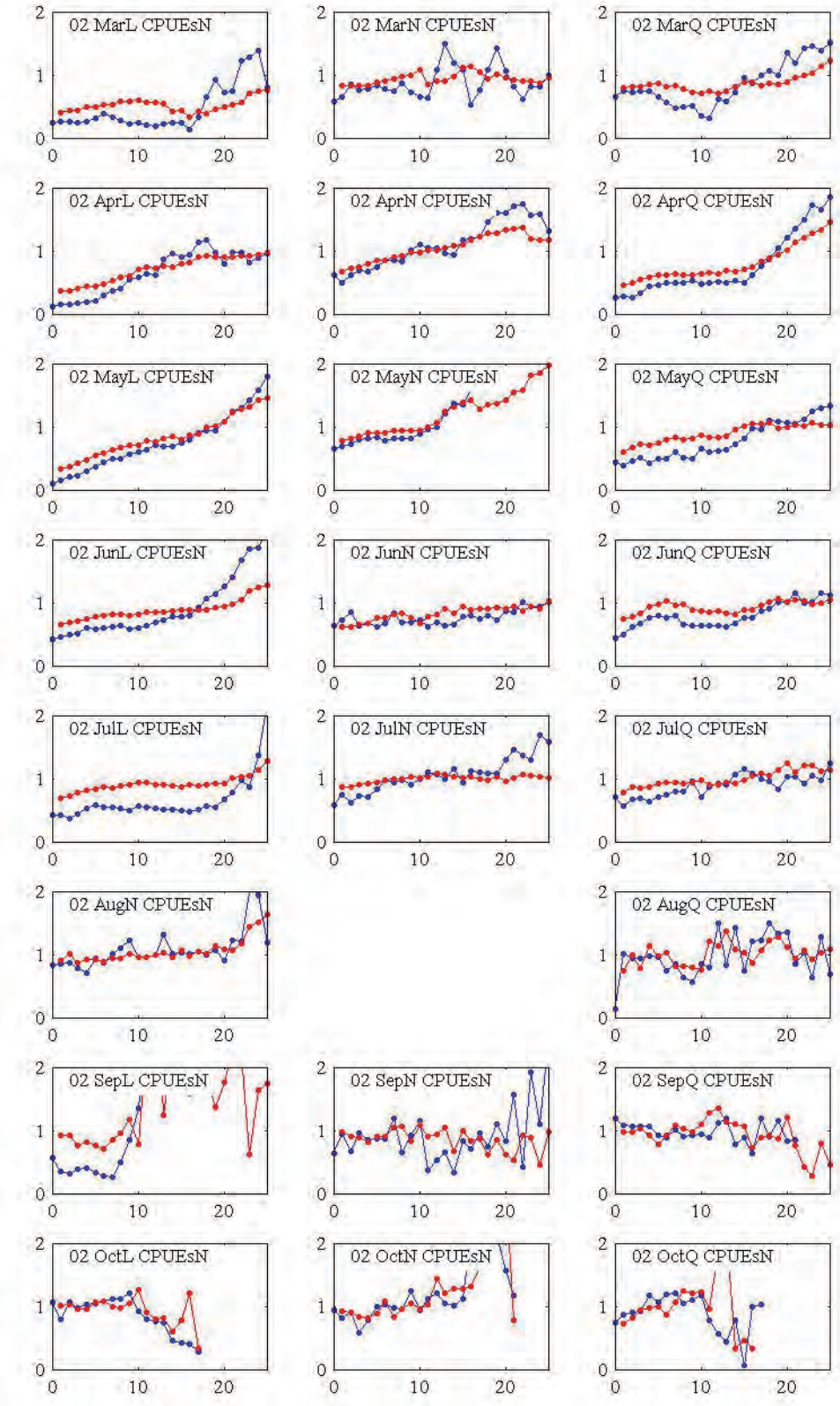


Figure H.3-3. Experimental semivariograms of *CPUEsNvd* and *CPUEsNgd* for Season 2003

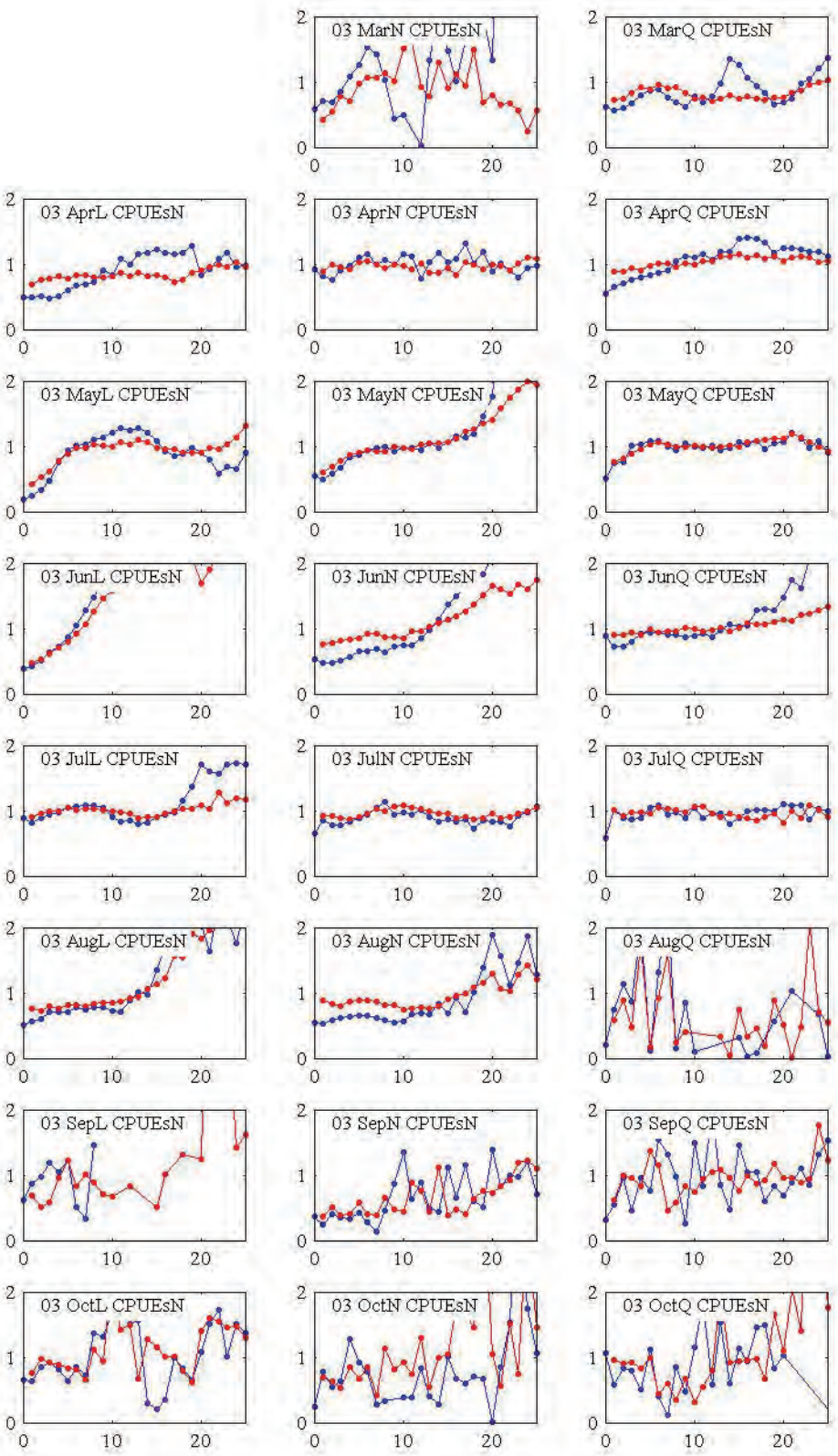


Figure H.3-4. Experimental semivariograms of *CPUEsNvd*, Seasons 2001-2004

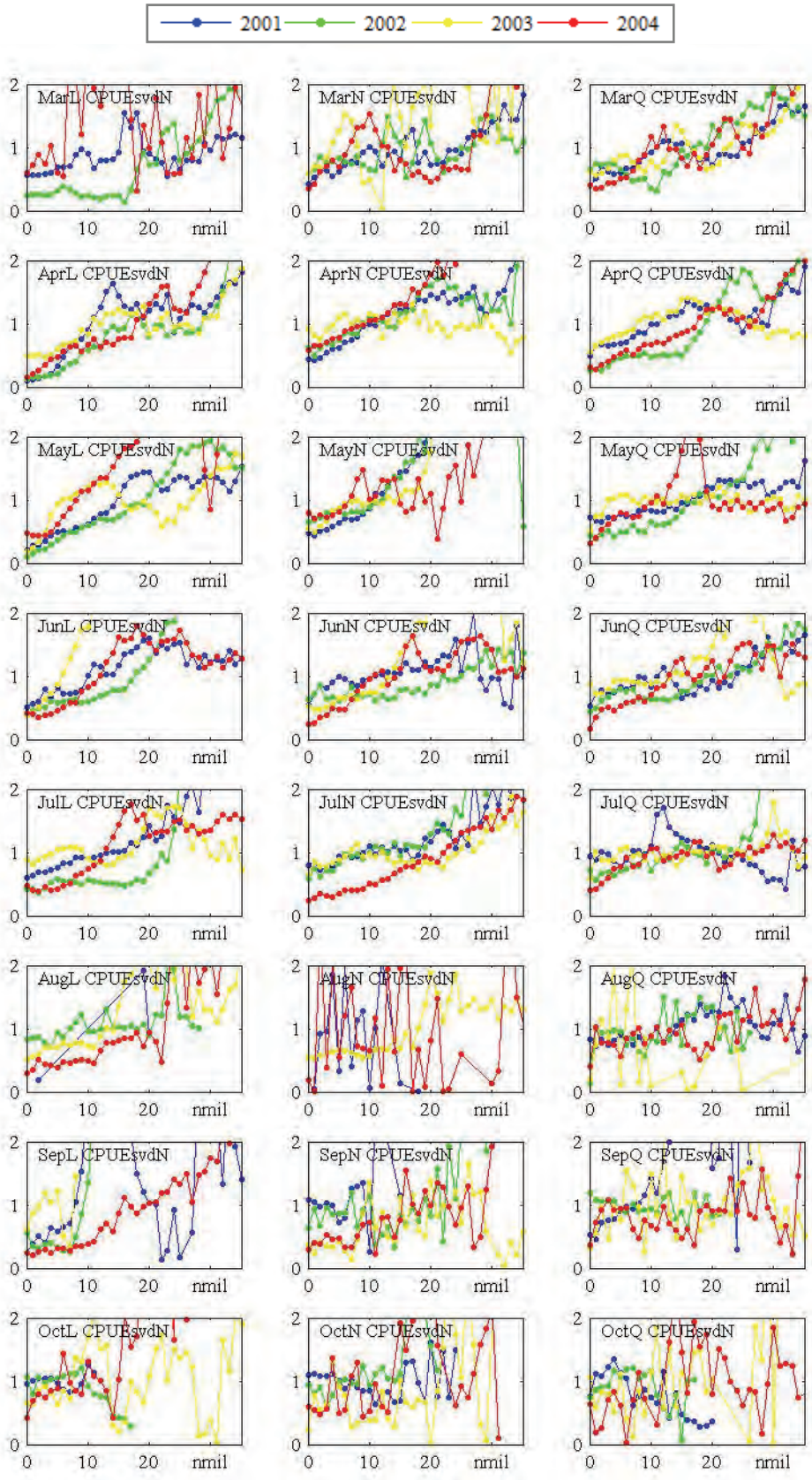
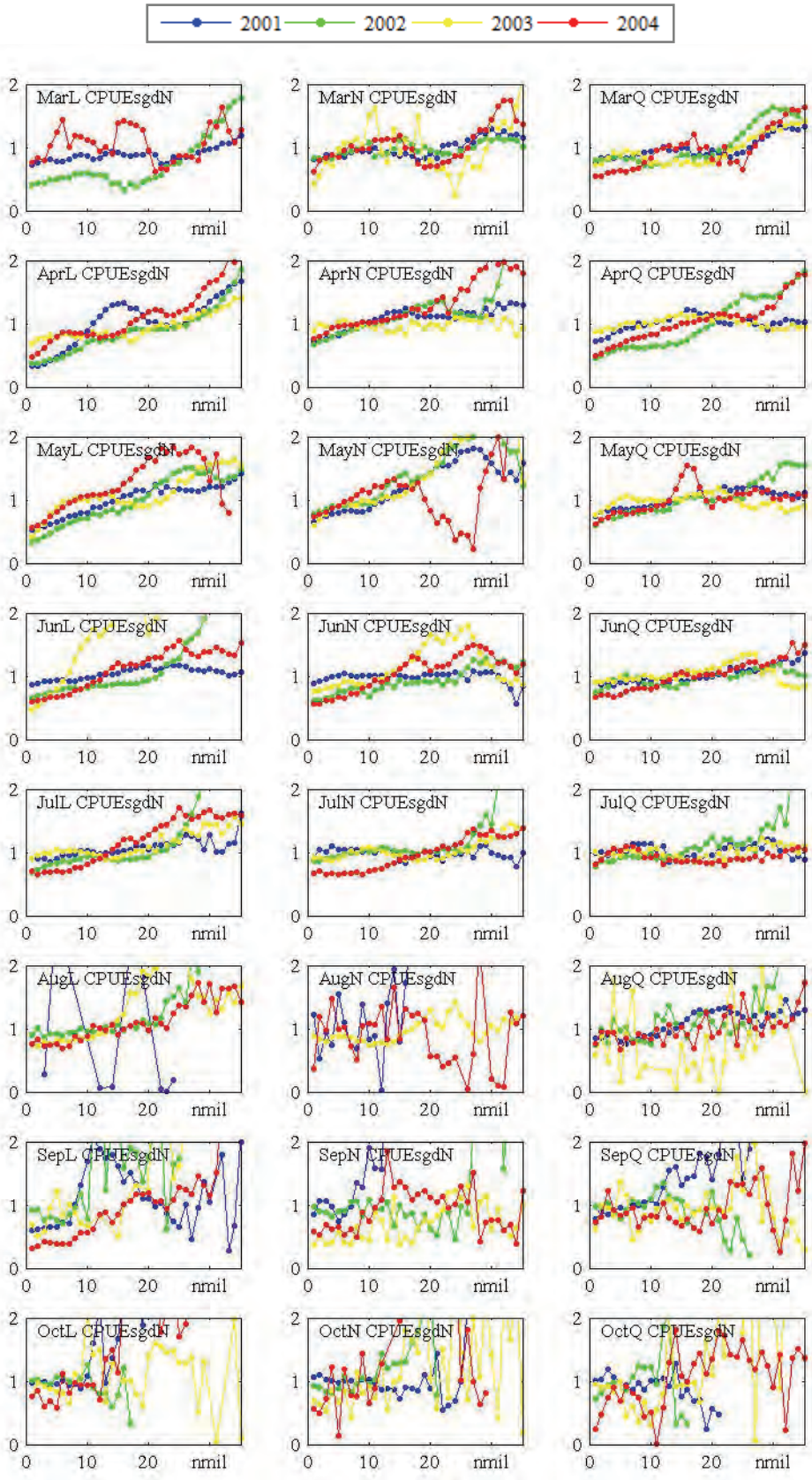


Figure H.3-5. Experimental semivariograms of *CPUEsNgd*, Seasons 2001-2004



APPENDIX I: SPATIOTEMPORAL VARIOGRAPHY OF COMBINED SEASONS

I.1 Spatiotemporal variography of combined seasons

Figure I.1-1. Experimental marginal spatial and temporal semivariograms, combined seasons, 2001-2004.

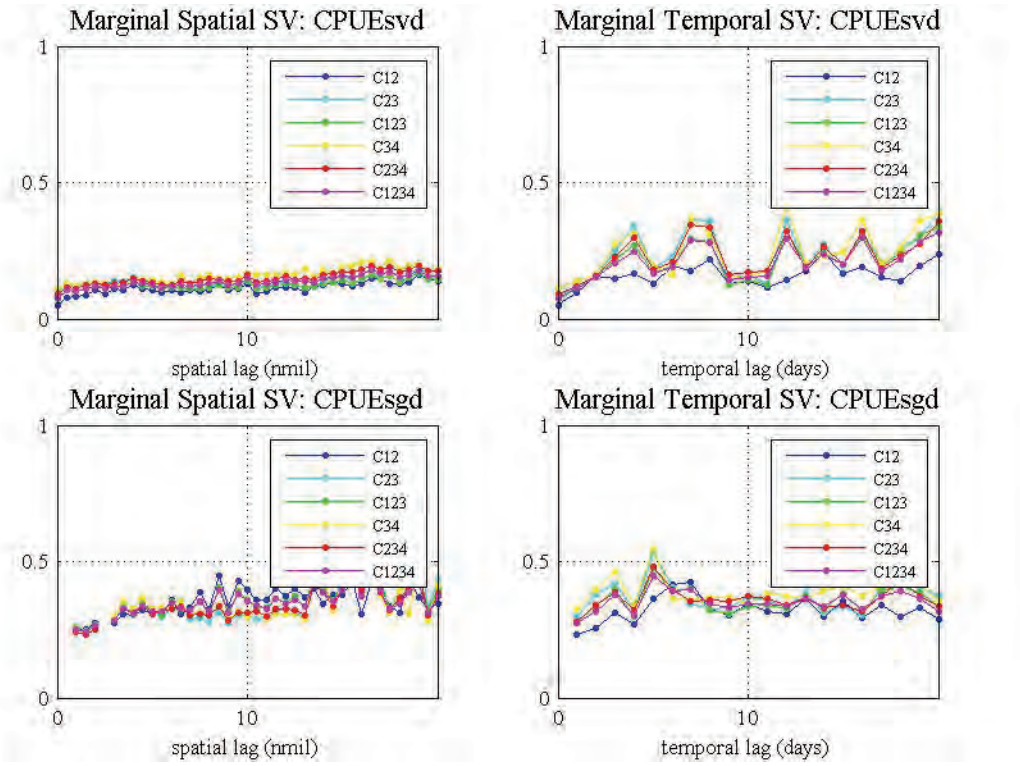


Figure I.1-2. Number of pairs for experimental marginal spatial and temporal semivariograms, combined seasons, 2001-2004.

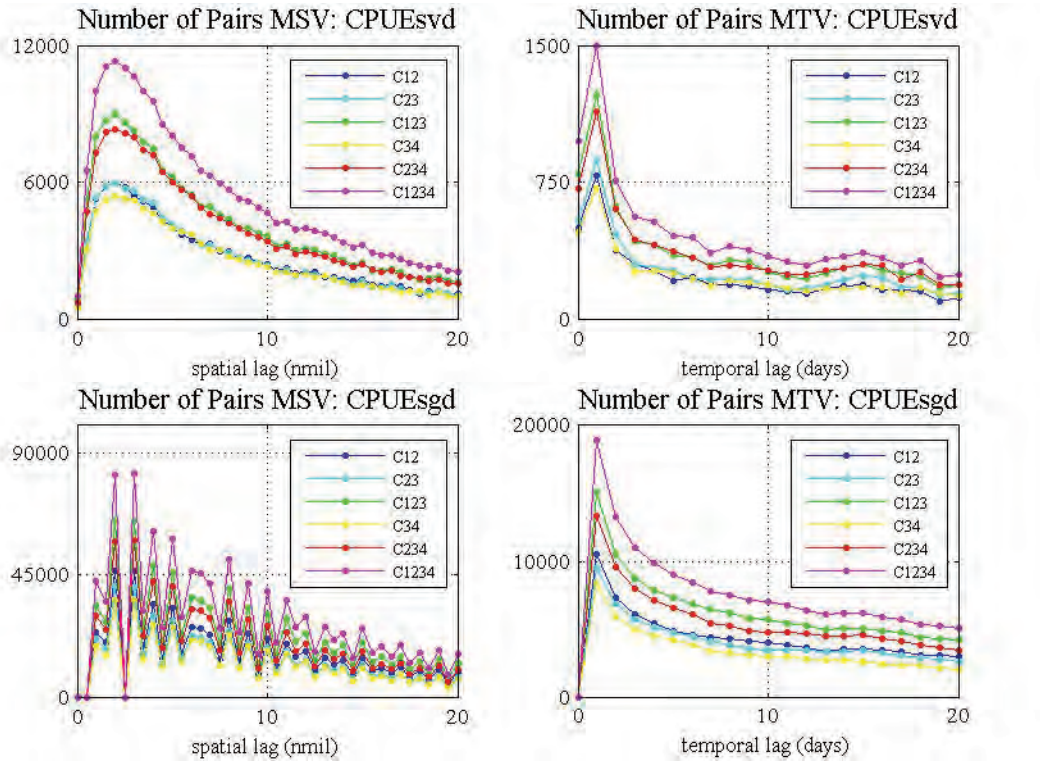
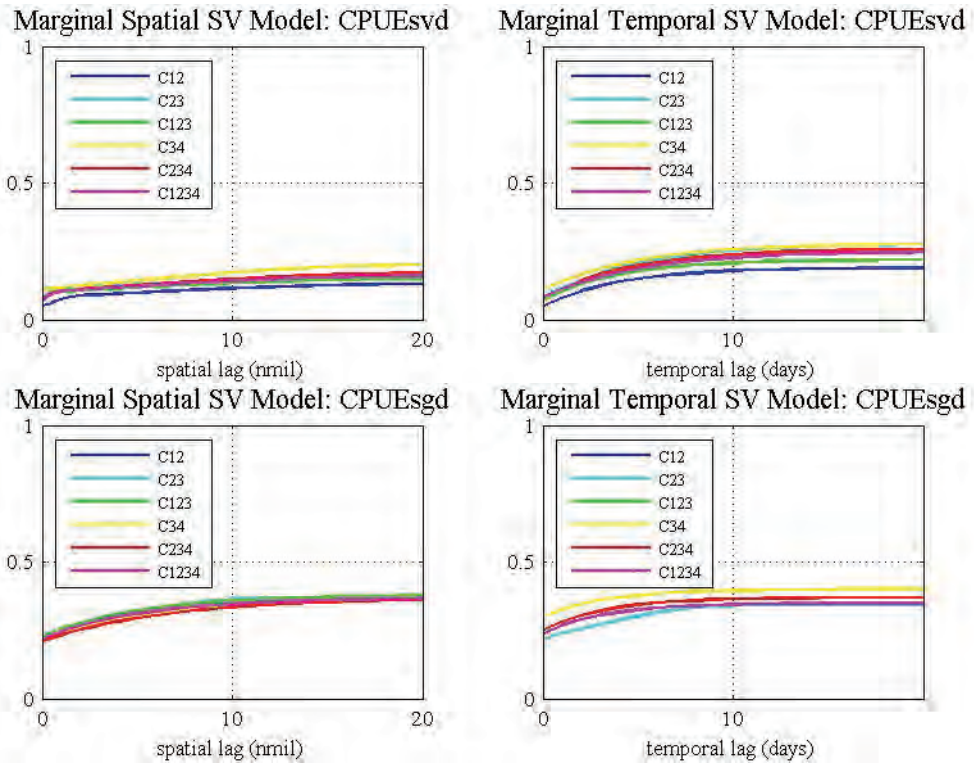
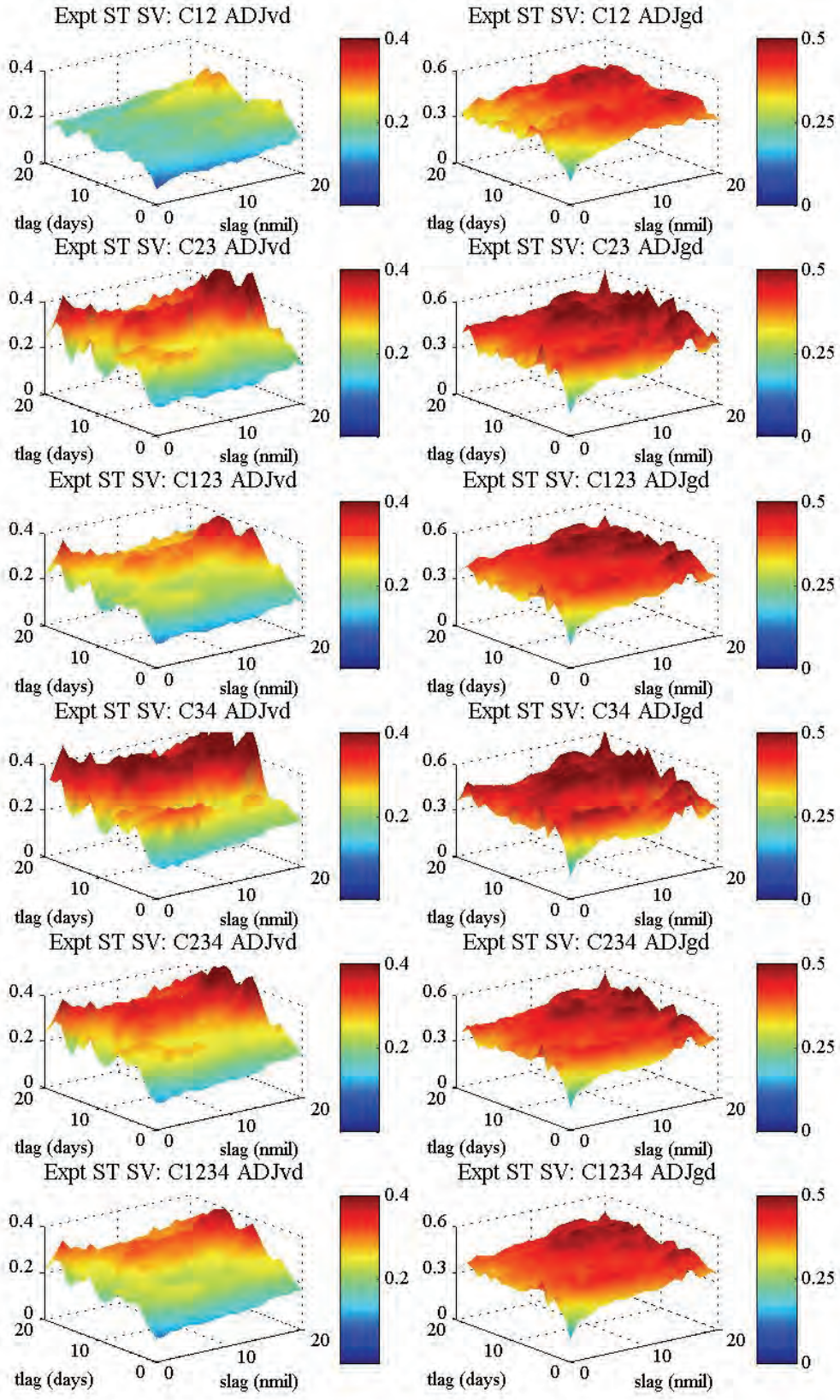


Figure I.1-3. Marginal spatial and temporal semivariogram models, combined seasons, 2001-2004.



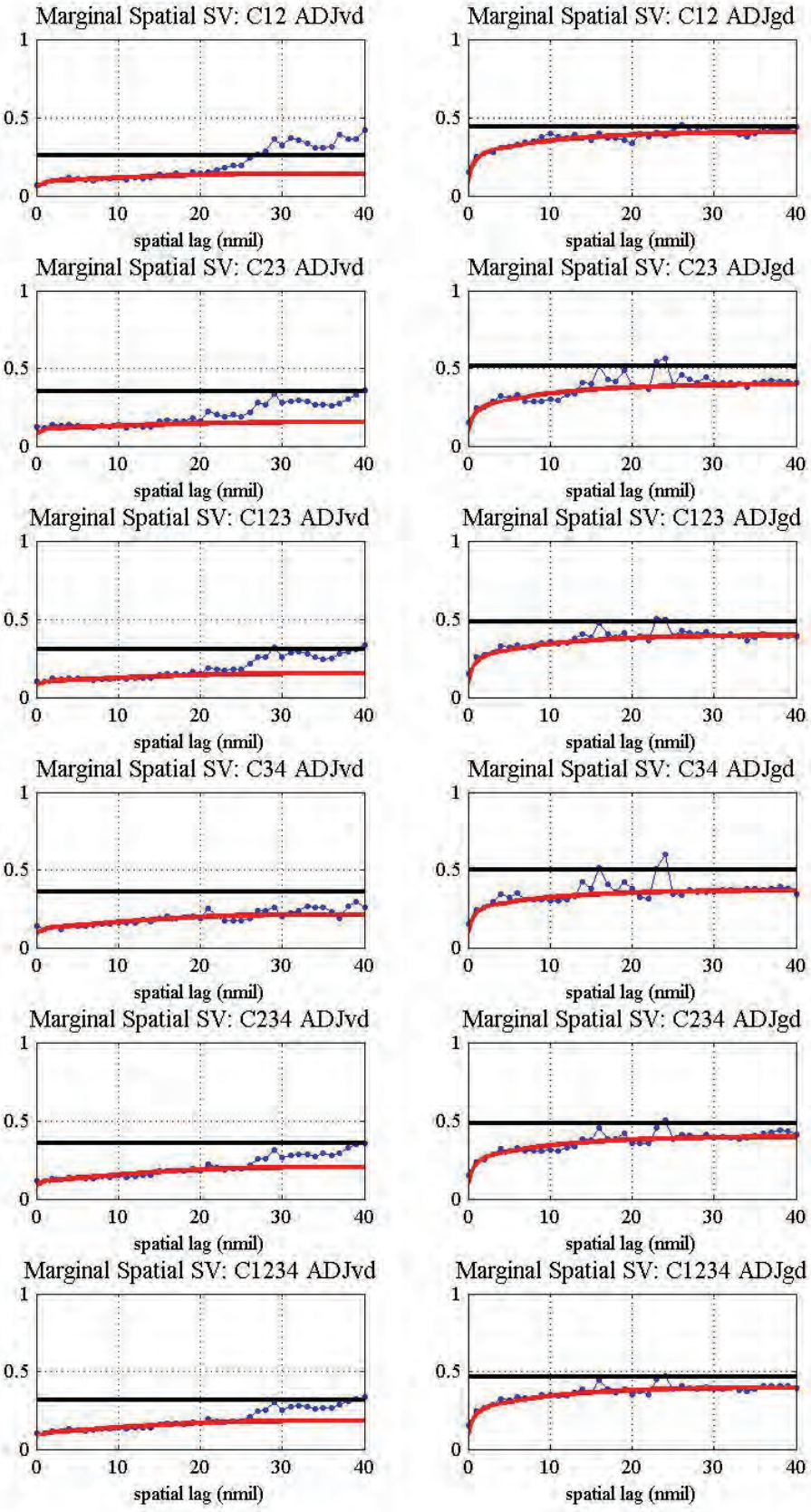
I.2 Experimental spatiotemporal semivariograms of combined seasons

Figure I.2-1. Experimental semivariograms, combined seasons, 2001-2004.



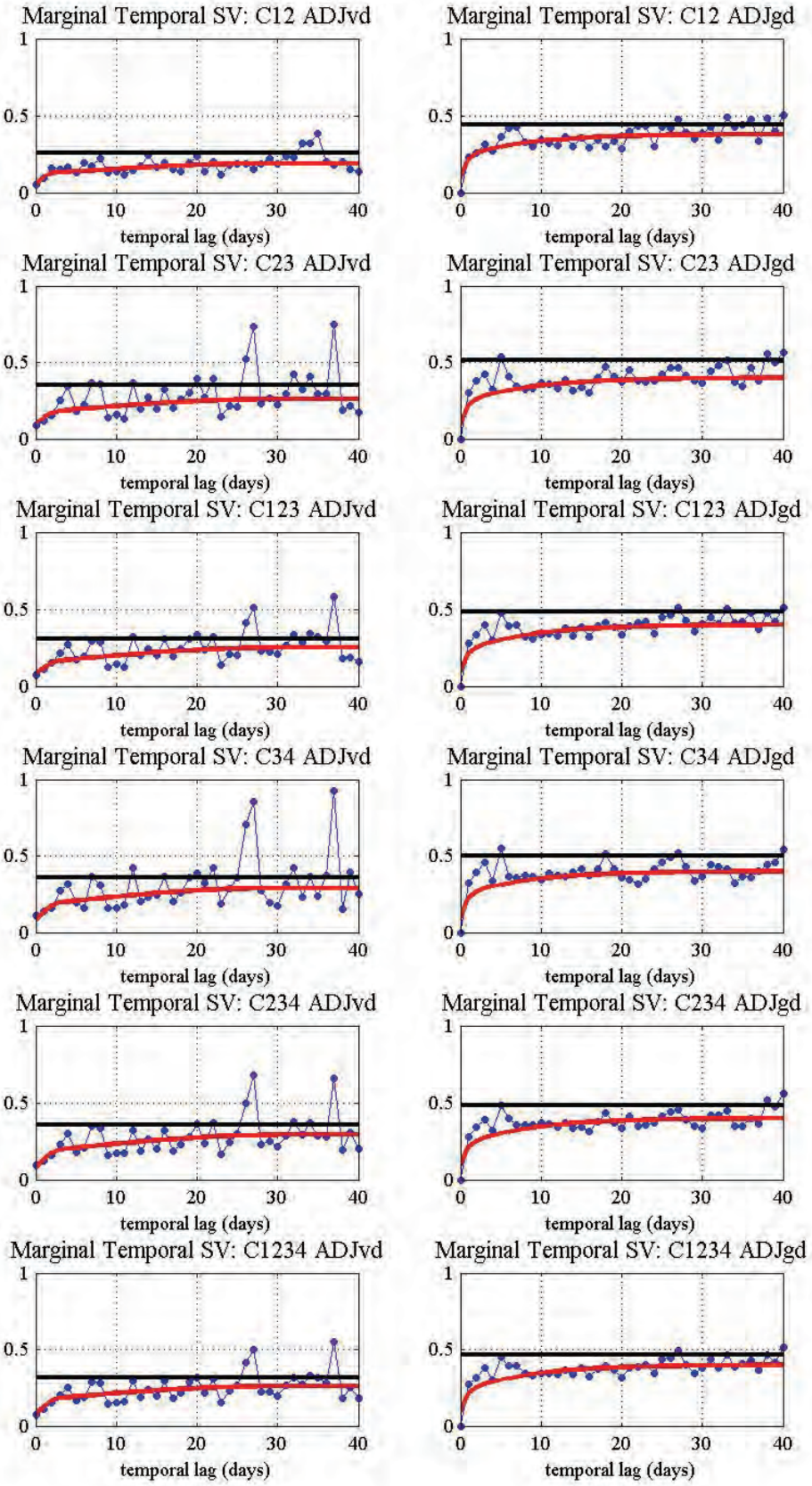
I.3 Marginal spatial semivariograms of combined seasons

Figure I.3-1. Marginal spatial semivariograms, combined seasons, 2001-2004.



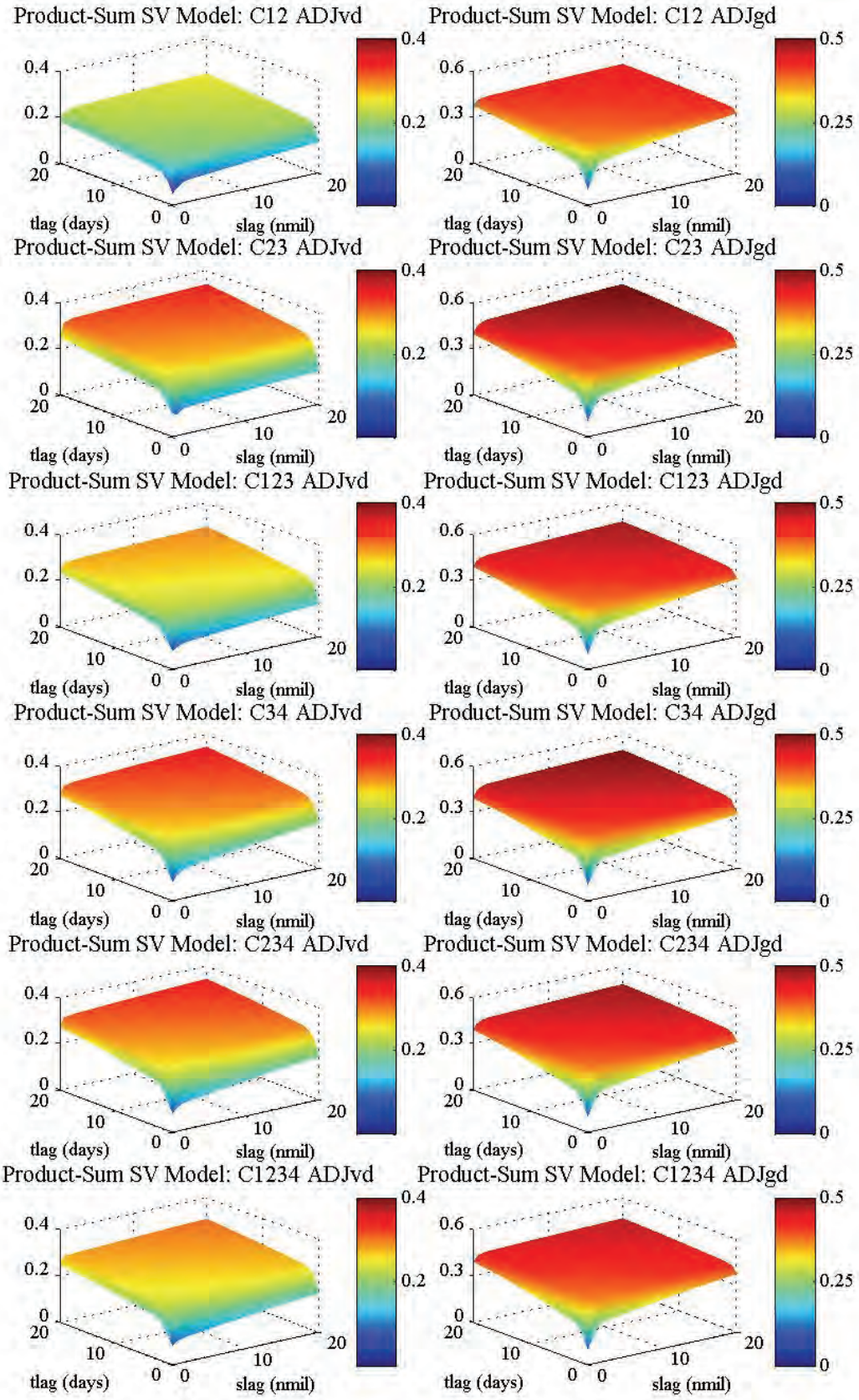
I.4 Marginal temporal semivariograms of combined seasons

Figure I.4-1. Marginal temporal semivariograms, combined seasons, 2001-2004.



I.5 Product-sum semivariogram models of combined seasons

Figure I.5-1. Product-sum semivariogram models, combined seasons, 2001-2004.



APPENDIX J: CROSS VALIDATION

J.1 Standardised catch rate data of 2004

Figure J.1-1. Standardised CPUE, March-June 2004

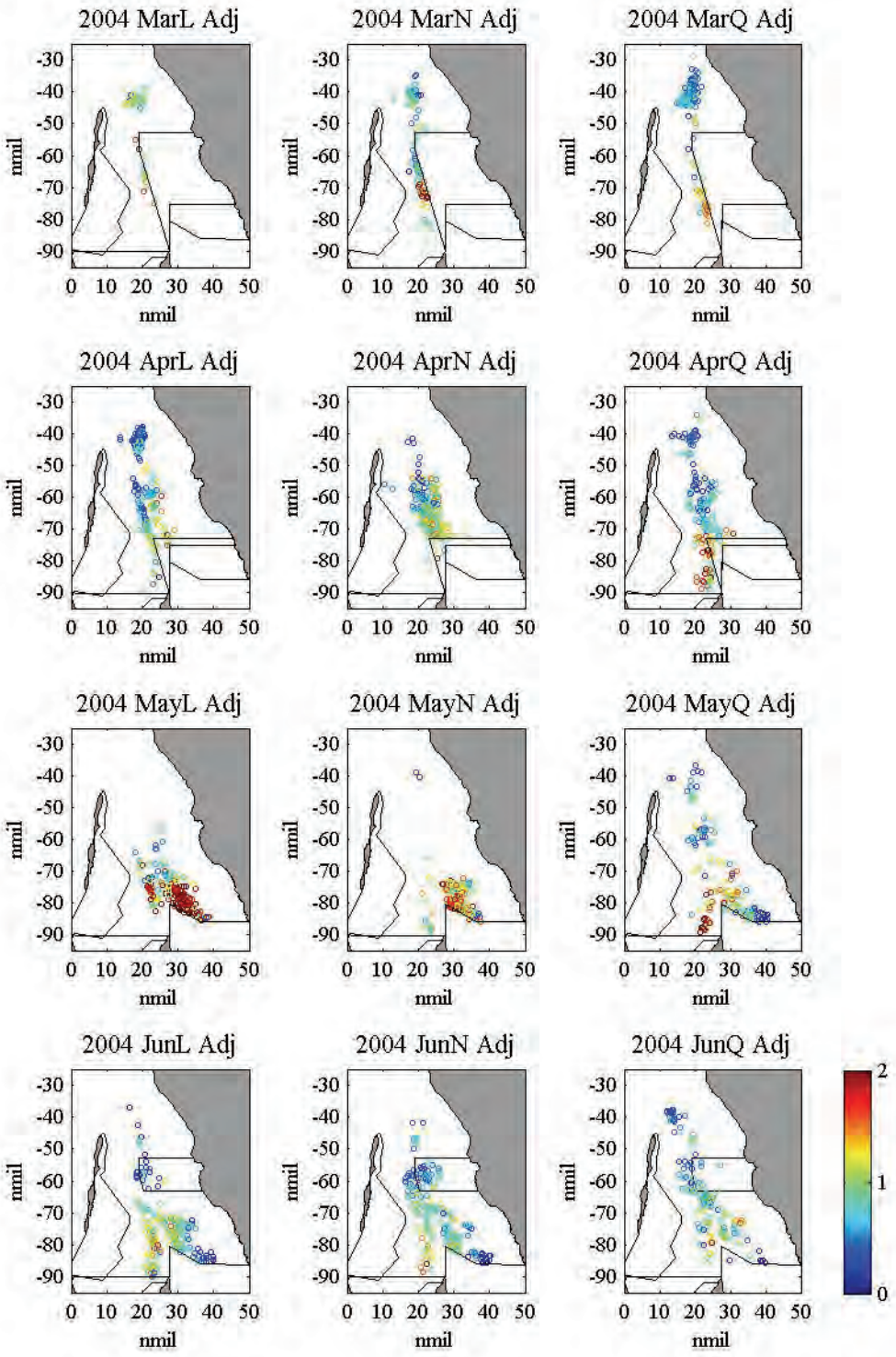
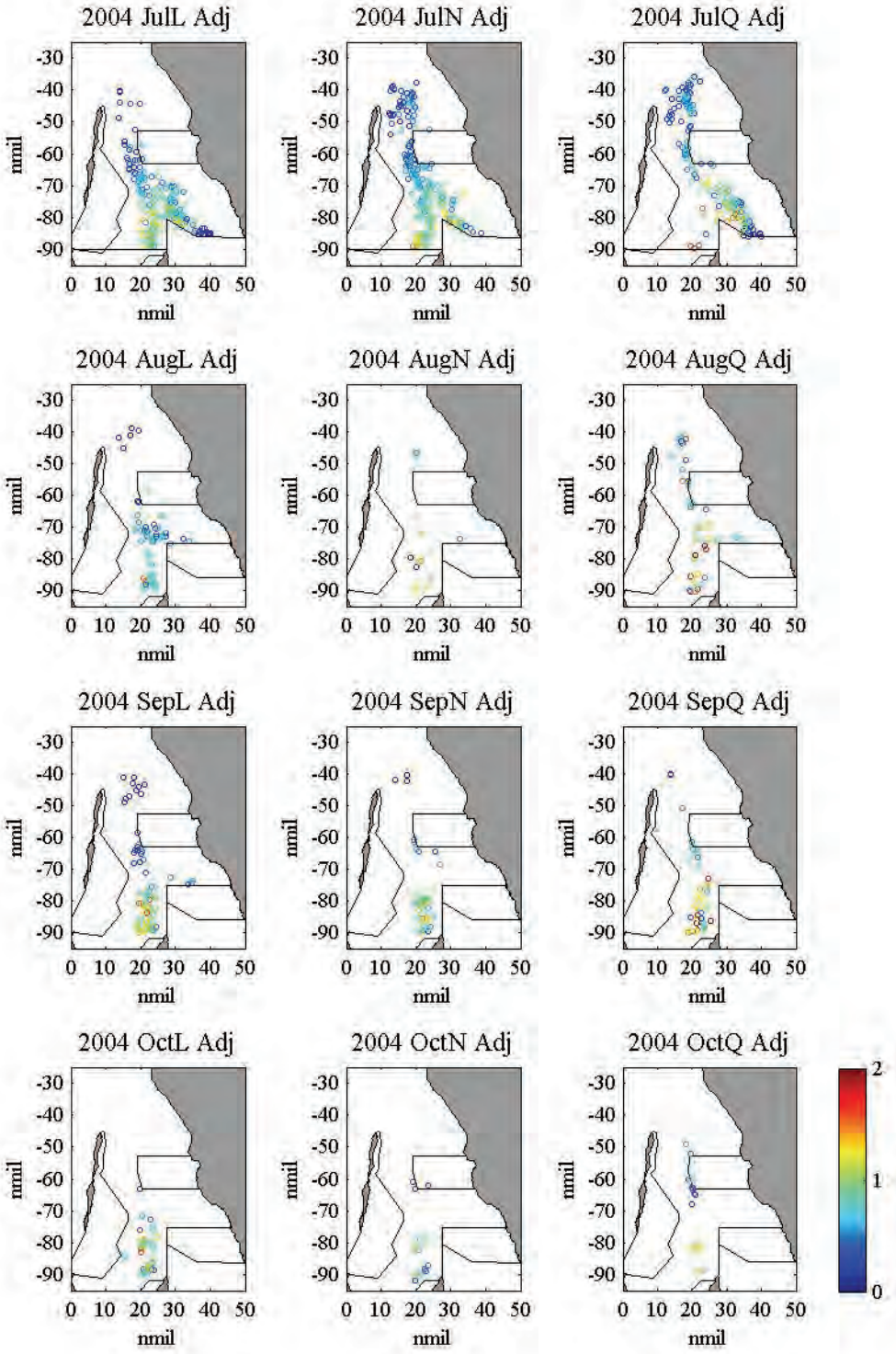


Figure J.1-2. Standardised CPUE, July-October 2004



J.2 Standardised CPUE Cross Validation Estimates 2004

Figure J.2-1. Cross validation estimates of CPUEs, March-June 2004.

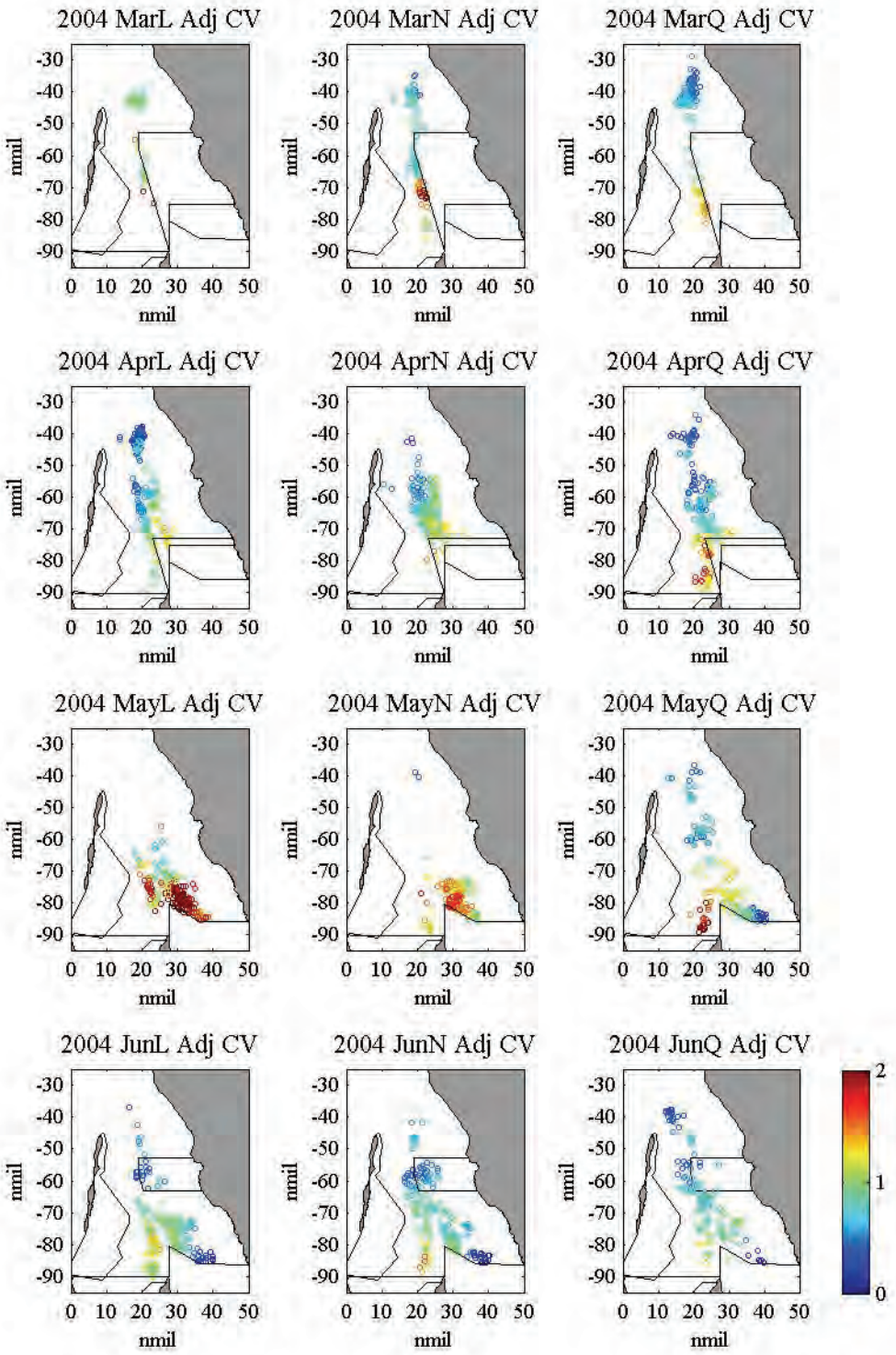
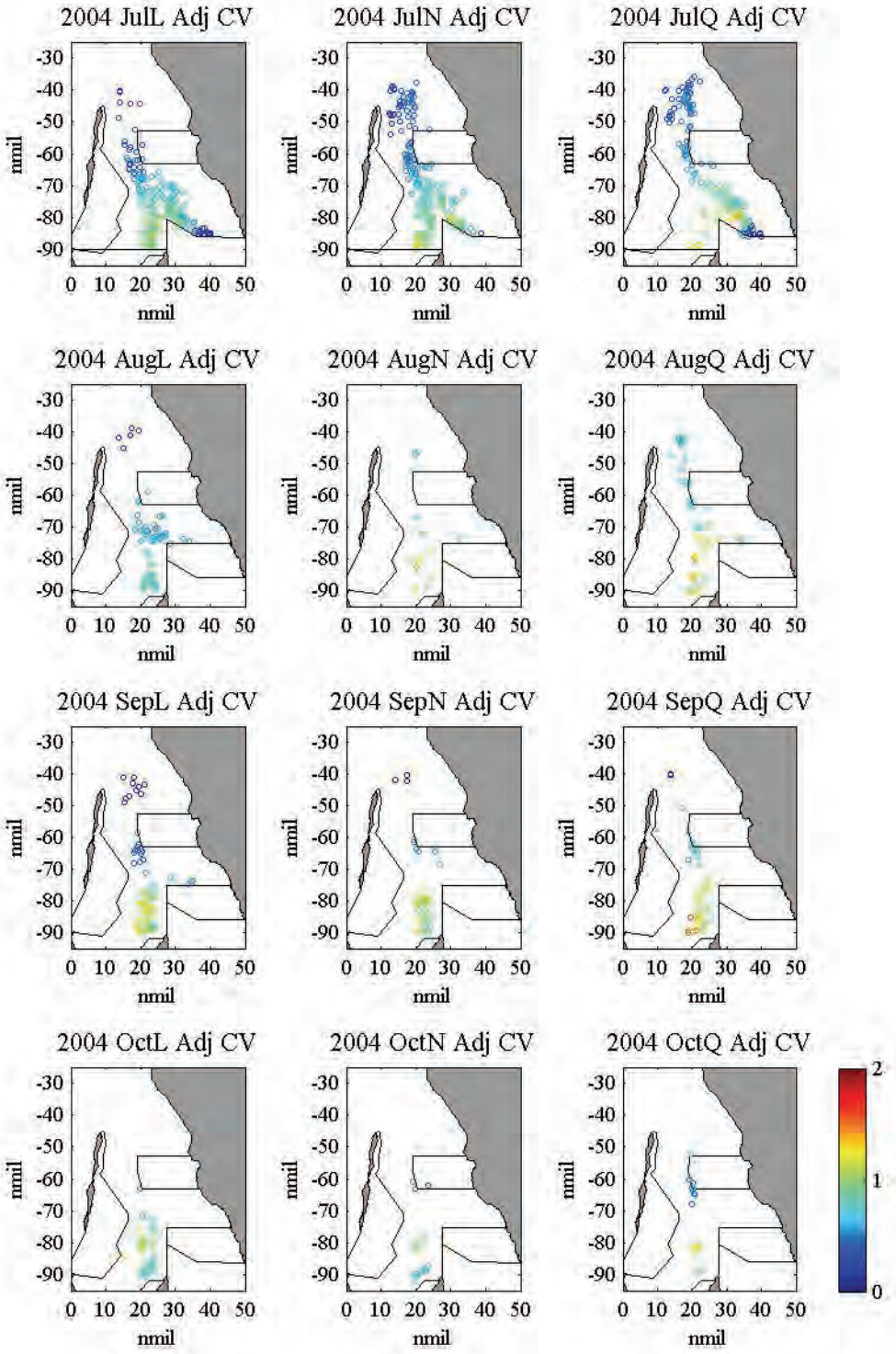


Figure J.2-2. Cross validation estimates of CPUEs, July-October 2004.



J.3 Summary statistics of standardised CPUE 2004

Table J.3-1. Summary statistics for standardised CPUE, 2004.

Week	Mean	Med	Var	Skew	Min	Max	N
MarL	1.026	1.034	0.130	0.645	0.348	2.222	52
MarN	0.942	0.815	0.267	1.512	0.063	3.223	112
MarQ	0.672	0.563	0.129	1.086	0.072	1.744	140
AprL	0.650	0.557	0.144	0.759	0.072	1.823	200
AprN	0.852	0.836	0.114	0.257	0.187	1.854	202
AprQ	0.840	0.752	0.255	0.913	0.087	2.373	178
MayL	1.652	1.745	0.455	-0.006	0.207	3.527	196
MayN	1.228	1.213	0.152	0.112	0.188	2.357	203
MayQ	0.917	0.844	0.382	1.285	0.036	3.271	165
JunL	0.824	0.838	0.138	-0.060	0.055	1.763	198
JunN	0.726	0.725	0.139	0.659	0.030	2.611	195
JunQ	0.729	0.729	0.121	0.394	0.047	1.892	143
JulL	0.694	0.677	0.091	0.043	0.038	1.459	197
JulN	0.700	0.684	0.086	0.316	0.150	1.495	202
JulQ	0.659	0.564	0.136	0.818	0.063	1.668	166
AugL	0.656	0.700	0.075	0.535	0.066	1.543	69
AugN	1.014	1.039	0.207	0.756	0.281	2.122	14
AugQ	1.000	0.920	0.215	0.706	0.330	2.213	48
SepL	0.797	0.844	0.191	0.093	0.100	1.701	74
SepN	0.852	0.836	0.134	0.103	0.065	1.788	45
SepQ	1.062	1.056	0.240	0.359	0.193	2.271	43
OctL	0.883	0.874	0.108	0.225	0.249	1.650	42
OctN	0.689	0.665	0.090	0.469	0.257	1.322	19
OctQ	0.745	0.547	0.183	0.690	0.225	1.636	25

J.4 Standardised CPUE cross validation summary statistics 2004

Table J.4-1. Summary statistics for standardised CPUE cross validation estimates, 2004.

Week	Mean	Med	Var	Skew	Min	Max	N
MarL	1.030	0.977	0.037	3.042	0.839	1.836	51
MarN	0.964	0.782	0.182	1.266	0.415	2.252	112
MarQ	0.678	0.585	0.092	1.272	0.255	1.520	140
AprL	0.650	0.627	0.096	0.304	0.203	1.382	200
AprN	0.852	0.888	0.052	0.014	0.361	1.563	202
AprQ	0.821	0.709	0.169	0.623	0.219	1.789	178
MayL	1.627	1.746	0.221	-0.477	0.577	2.557	196
MayN	1.265	1.232	0.065	-0.086	0.390	1.796	203
MayQ	0.911	0.810	0.219	0.927	0.222	2.243	165
JunL	0.833	0.911	0.081	-0.574	0.130	1.433	198
JunN	0.724	0.749	0.088	0.049	0.072	1.635	195
JunQ	0.710	0.773	0.064	-0.288	0.099	1.241	143
JulL	0.704	0.723	0.055	-0.385	0.164	1.095	197
JulN	0.707	0.721	0.060	0.045	0.244	1.248	202
JulQ	0.654	0.573	0.068	0.562	0.237	1.321	166
AugL	0.674	0.637	0.031	-0.299	0.228	1.000	69
AugN	0.971	0.932	0.052	0.286	0.597	1.426	14
AugQ	0.967	0.962	0.057	0.187	0.529	1.402	48
SepL	0.816	0.972	0.133	-0.562	0.164	1.303	74
SepN	0.916	1.020	0.077	-1.398	0.171	1.245	45
SepQ	1.015	1.032	0.108	-0.385	0.179	1.621	43
OctL	0.898	0.861	0.025	-0.065	0.494	1.162	42
OctN	0.770	0.721	0.044	0.045	0.393	1.110	19
OctQ	0.717	0.665	0.086	0.689	0.376	1.269	24

Table J.4-2. Accuracy measures for standardised CPUE cross validation estimates, 2004.

Week	ME	MSE	MAD	MPE	MAPE	RSQ	MStdE	MSDR	r(True/ CVest)	r(CVest/ CVerr)
MarL	-0.003	0.105	0.249	12.740	31.171	0.184	-0.002	0.655	0.429	0.109
MarN	0.022	0.093	0.226	20.087	36.048	0.652	0.047	0.563	0.808	0.019
MarQ	0.006	0.043	0.162	16.118	34.525	0.664	0.014	0.247	0.815	0.048
AprL	0.000	0.051	0.164	13.538	31.895	0.646	-0.004	0.294	0.804	0.020
AprN	0.000	0.082	0.218	13.090	31.581	0.296	-0.003	0.454	0.544	0.147
AprQ	-0.020	0.081	0.209	12.392	32.472	0.680	-0.046	0.459	0.825	-0.020
MayL	-0.025	0.221	0.340	12.218	28.397	0.514	-0.056	1.366	0.717	-0.030
MayN	0.037	0.099	0.235	14.487	26.857	0.361	0.086	0.576	0.601	0.067
MayQ	-0.006	0.172	0.294	40.130	61.844	0.546	-0.021	0.957	0.739	0.021
JunL	0.009	0.064	0.195	17.834	35.369	0.535	0.022	0.368	0.731	0.050
JunN	-0.003	0.055	0.175	21.346	41.135	0.605	-0.005	0.303	0.778	0.030
JunQ	-0.019	0.063	0.187	12.160	32.935	0.480	-0.047	0.331	0.693	0.042
JulL	0.010	0.041	0.164	18.863	35.976	0.552	0.025	0.236	0.743	0.051
JulN	0.008	0.031	0.136	9.614	24.765	0.637	0.018	0.168	0.798	0.064
JulQ	-0.005	0.057	0.188	20.803	41.668	0.581	-0.012	0.307	0.762	-0.090
AugL	0.018	0.049	0.157	20.529	34.801	0.343	0.042	0.295	0.586	0.076
AugN	-0.043	0.233	0.324	25.575	51.429	0.003	-0.086	0.973	0.054	0.407
AugQ	-0.033	0.173	0.336	14.363	39.048	0.191	-0.075	0.971	0.437	0.100
SepL	0.019	0.048	0.172	12.939	27.570	0.751	0.038	0.268	0.866	-0.061
SepN	0.064	0.075	0.227	20.050	33.910	0.466	0.139	0.410	0.683	0.101
SepQ	-0.048	0.216	0.346	17.287	45.308	0.160	-0.108	1.192	0.400	0.271
OctL	0.014	0.093	0.251	16.660	35.746	0.130	0.028	0.535	0.360	0.144
OctN	0.080	0.066	0.215	25.090	37.447	0.324	0.177	0.345	0.570	0.162
OctQ	-0.037	0.122	0.243	16.804	40.684	0.340	-0.084	0.637	0.583	0.116

J.5 Standardised CPUE cross validation estimate errors 2004

Figure J.5-1. Cross validation errors, March-June 2004. *Adj* denotes standardised catch rate *CPUEs*.

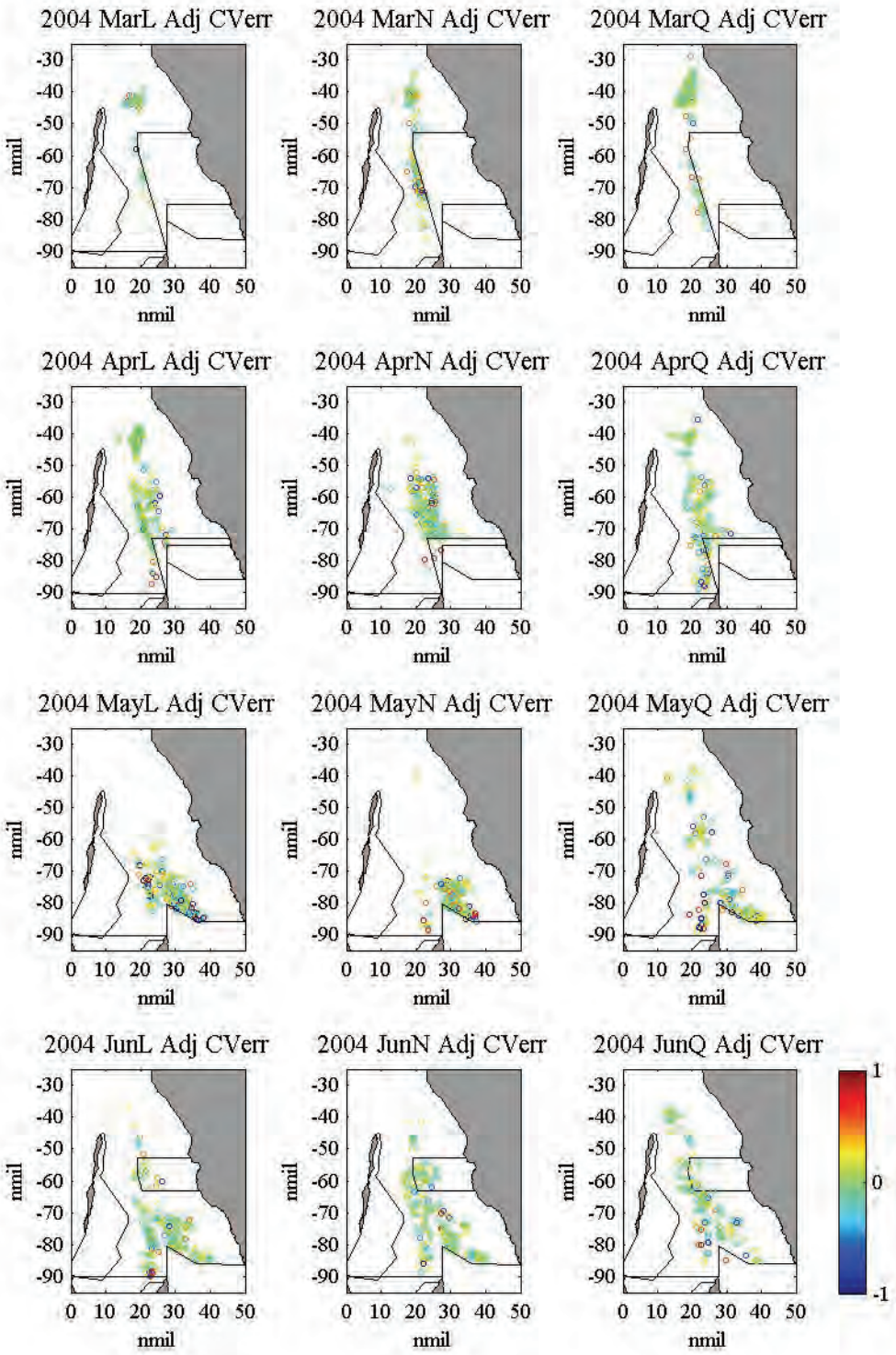
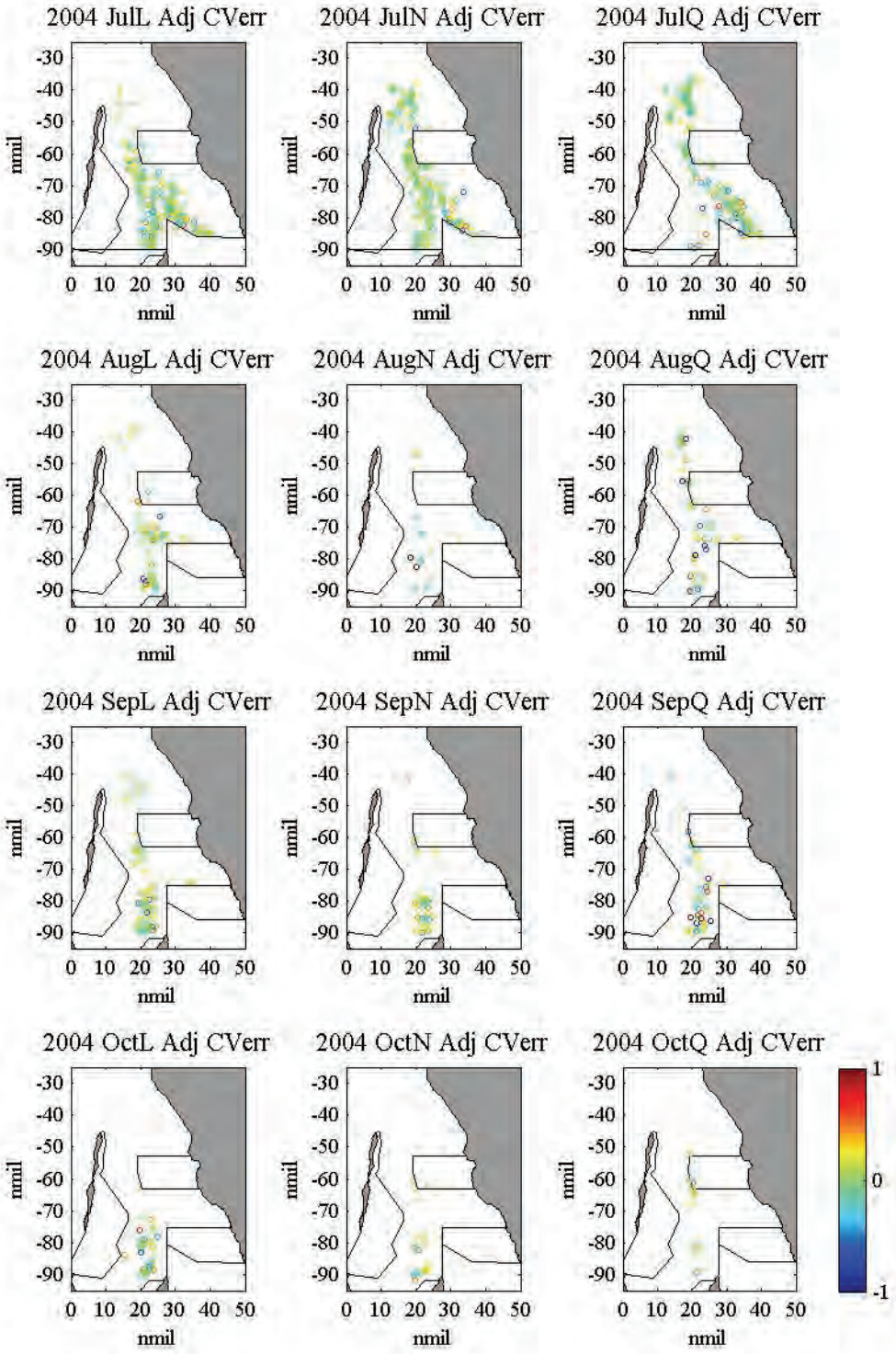


Figure J.5-2. Cross validation errors, July-October 2004. *Adj* denotes standardised catch rate *CPUEs*.



J.6 Histograms and scatterplots of standardised CPUE cross validation estimates 2004

Figure J.6-1. Scatterplots of CPUEs cross validation estimates, March-June 2004. *Adj* denotes standardised catch rate *CPUEs*.

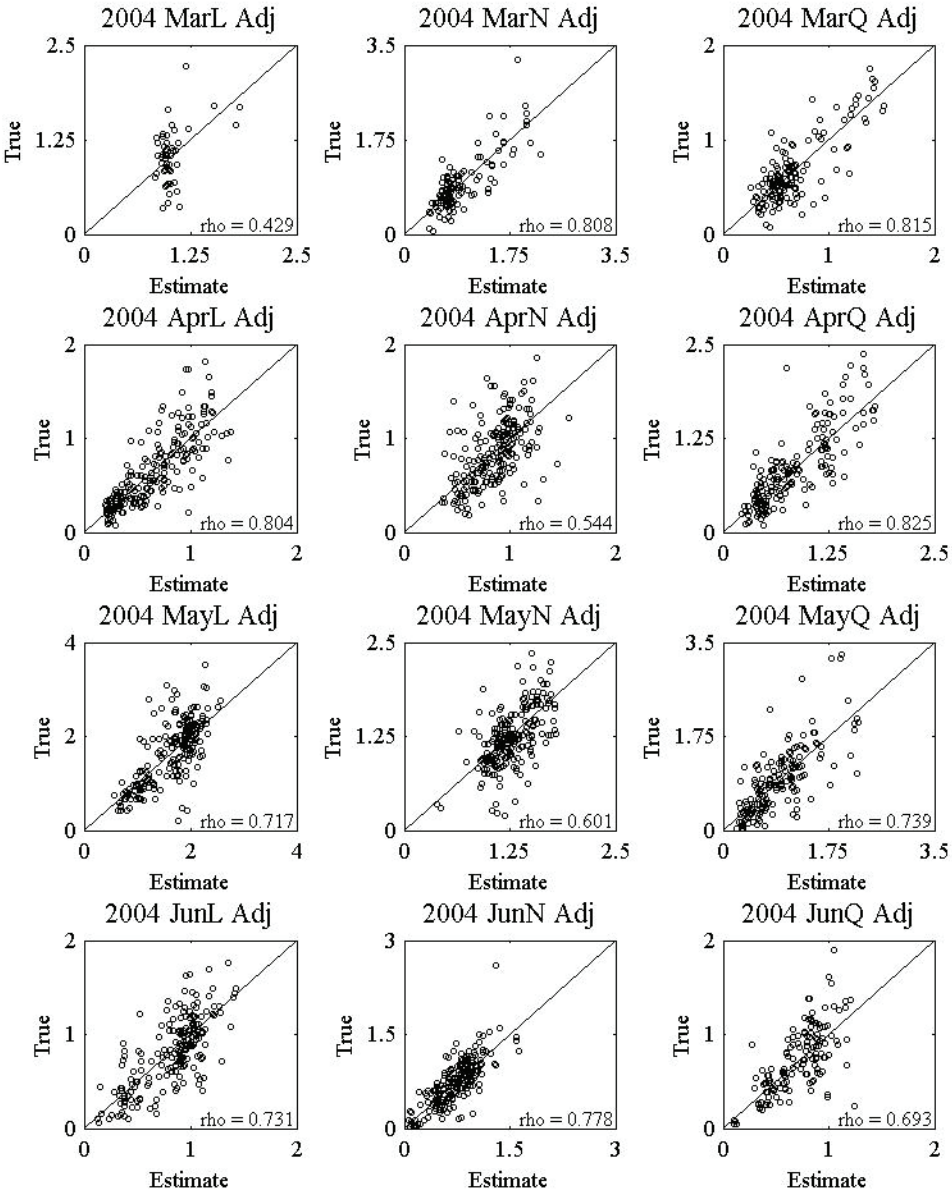


Figure J.6-2. Scatterplots of CPUEs cross validation estimates, July-October 2004. *Adj* denotes standardised catch rate *CPUEs*.

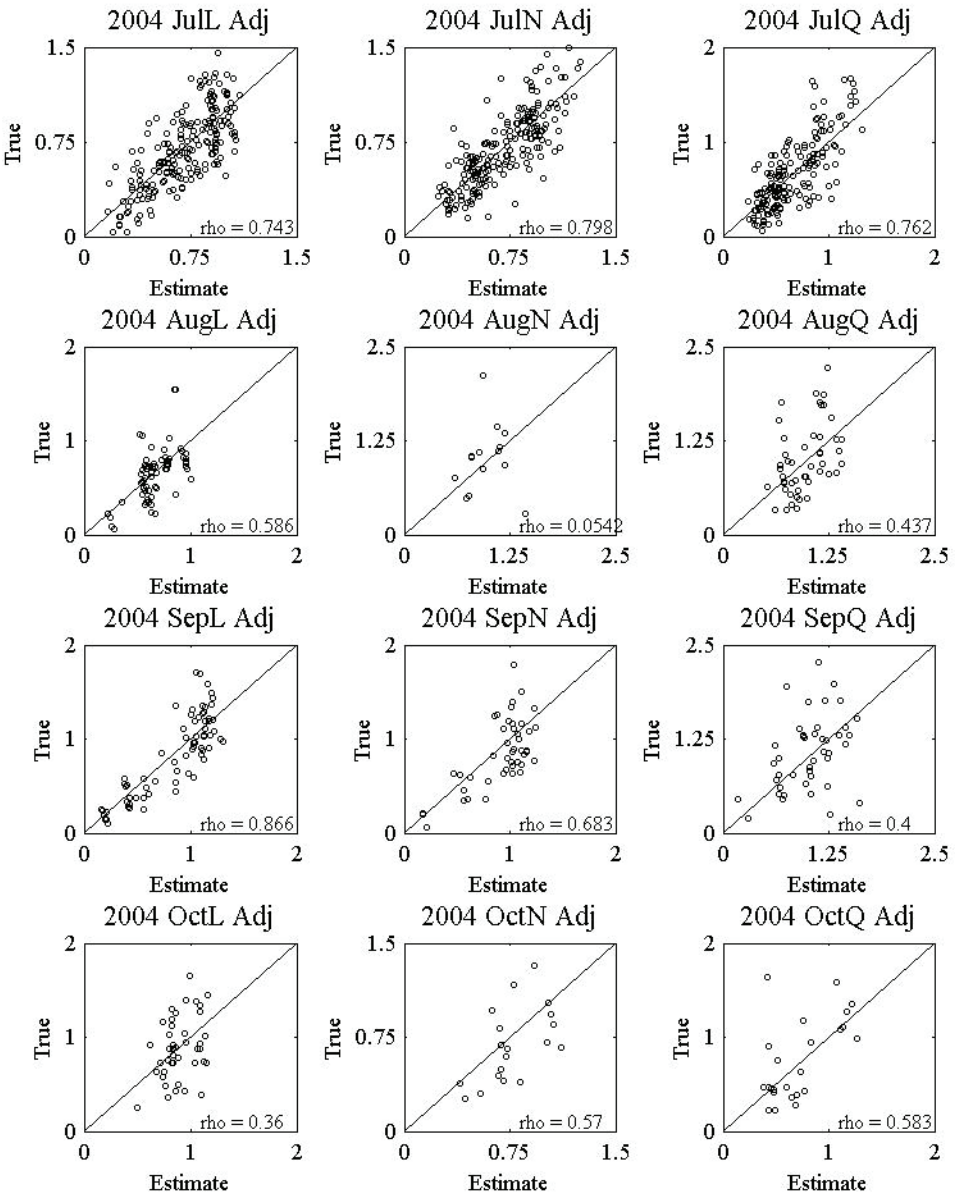


Figure J.6-3. Histograms of CPUEs cross validation estimates, March-June 2004. *Adj* denotes standardised catch rate *CPUEs*.

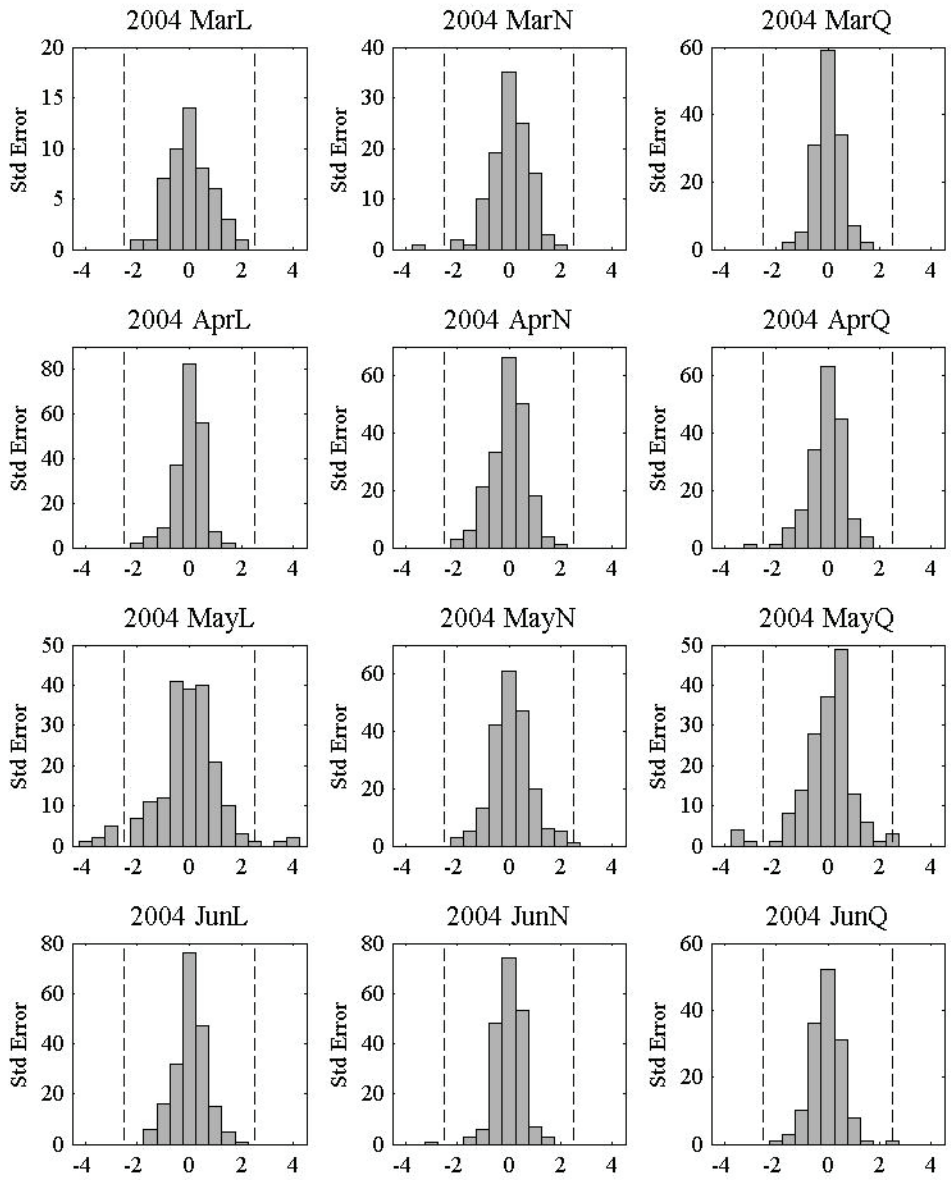


Figure J.6-4. Histograms of CPUEs cross validation estimates, July-October 2004.

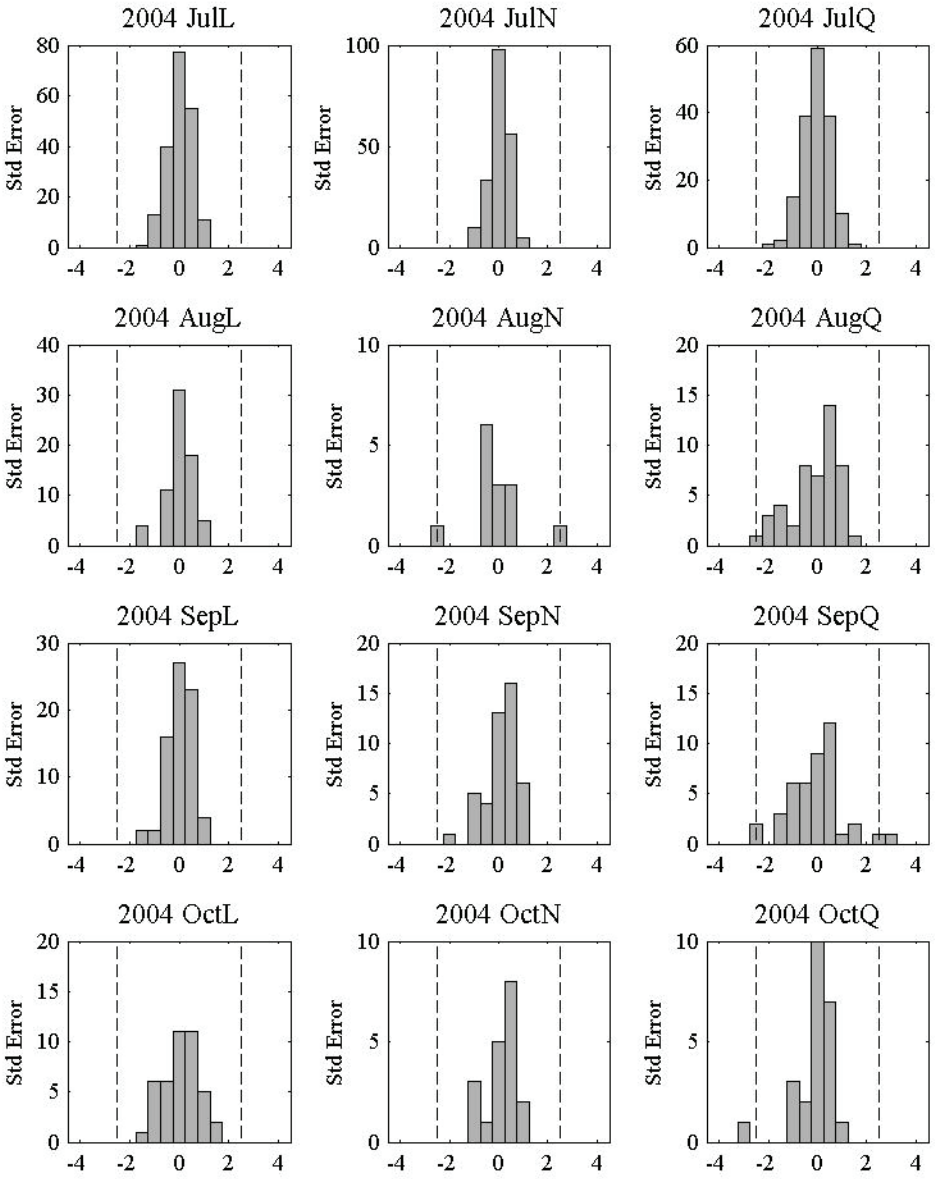


Figure J.6-5. Errors versus estimates of CPUEs cross validation estimates, March-June 2004. *Adj* denotes standardised catch rate *CPUEs*.

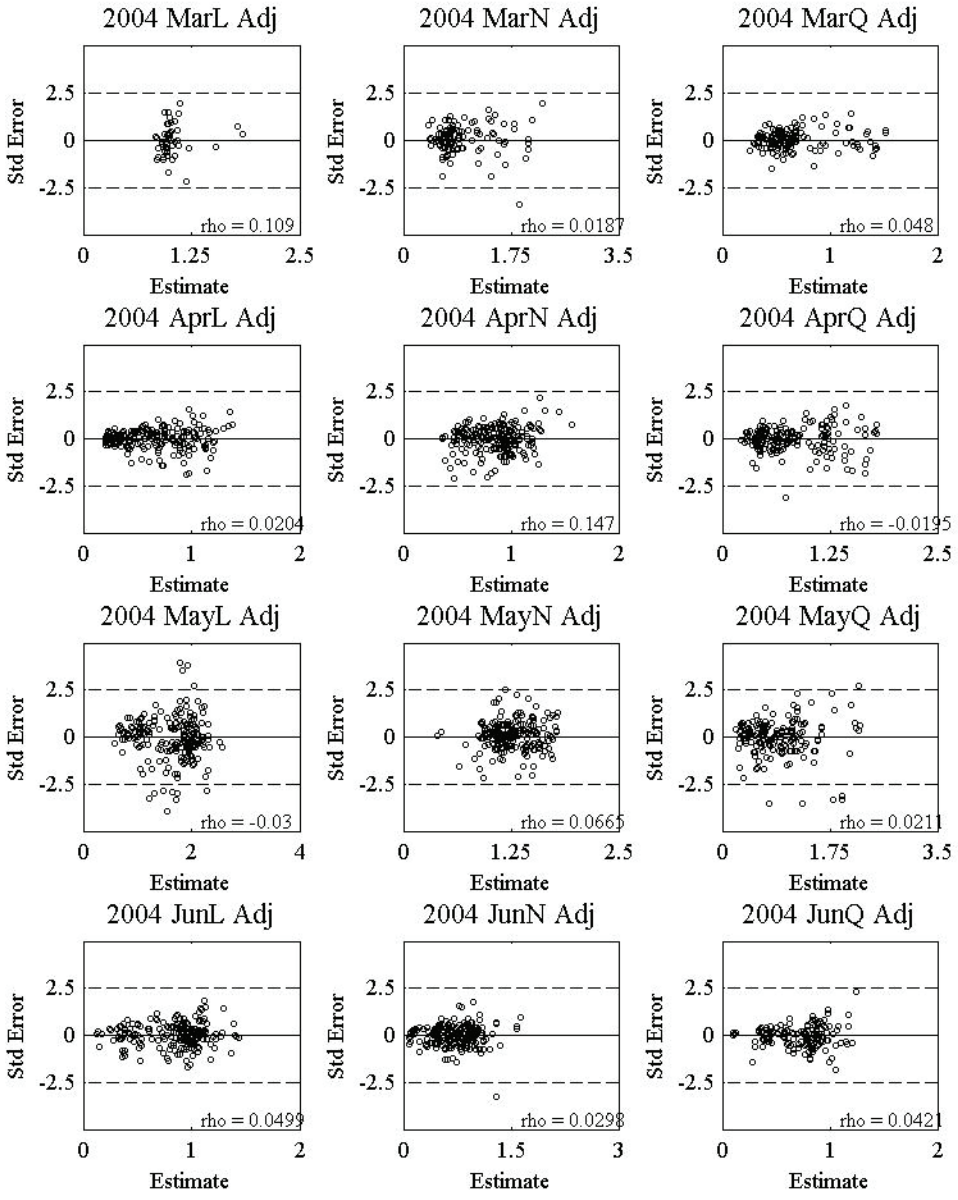
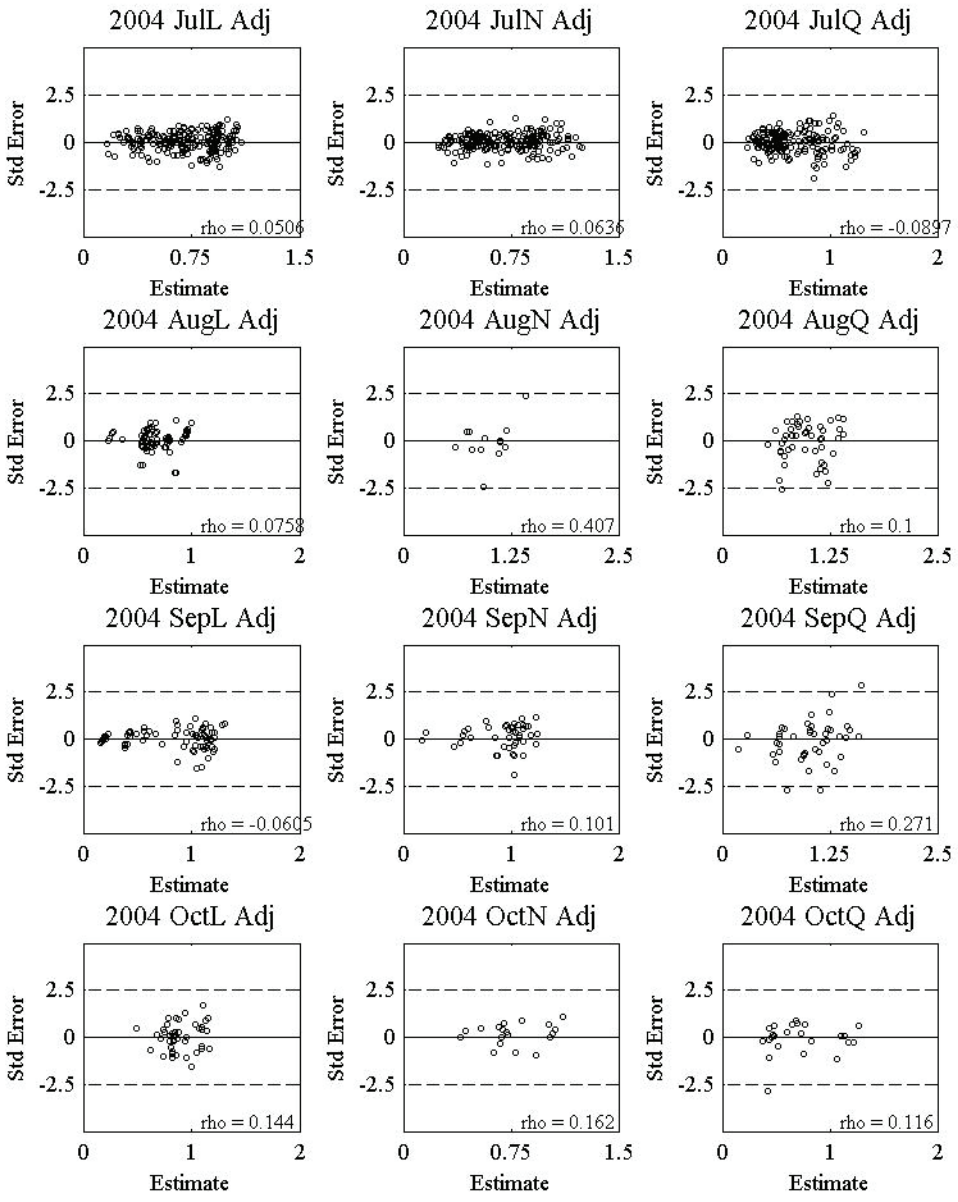


Figure J.6-6. Errors versus estimates of CPUEs cross validation estimates, July-October 2004. *Adj* denotes standardised catch rate *CPUEs*.



J.7 Cross Validation Estimates of CPUE 2004

Figure J.7-1. Cross validation estimates of CPUE, March-June 2004.

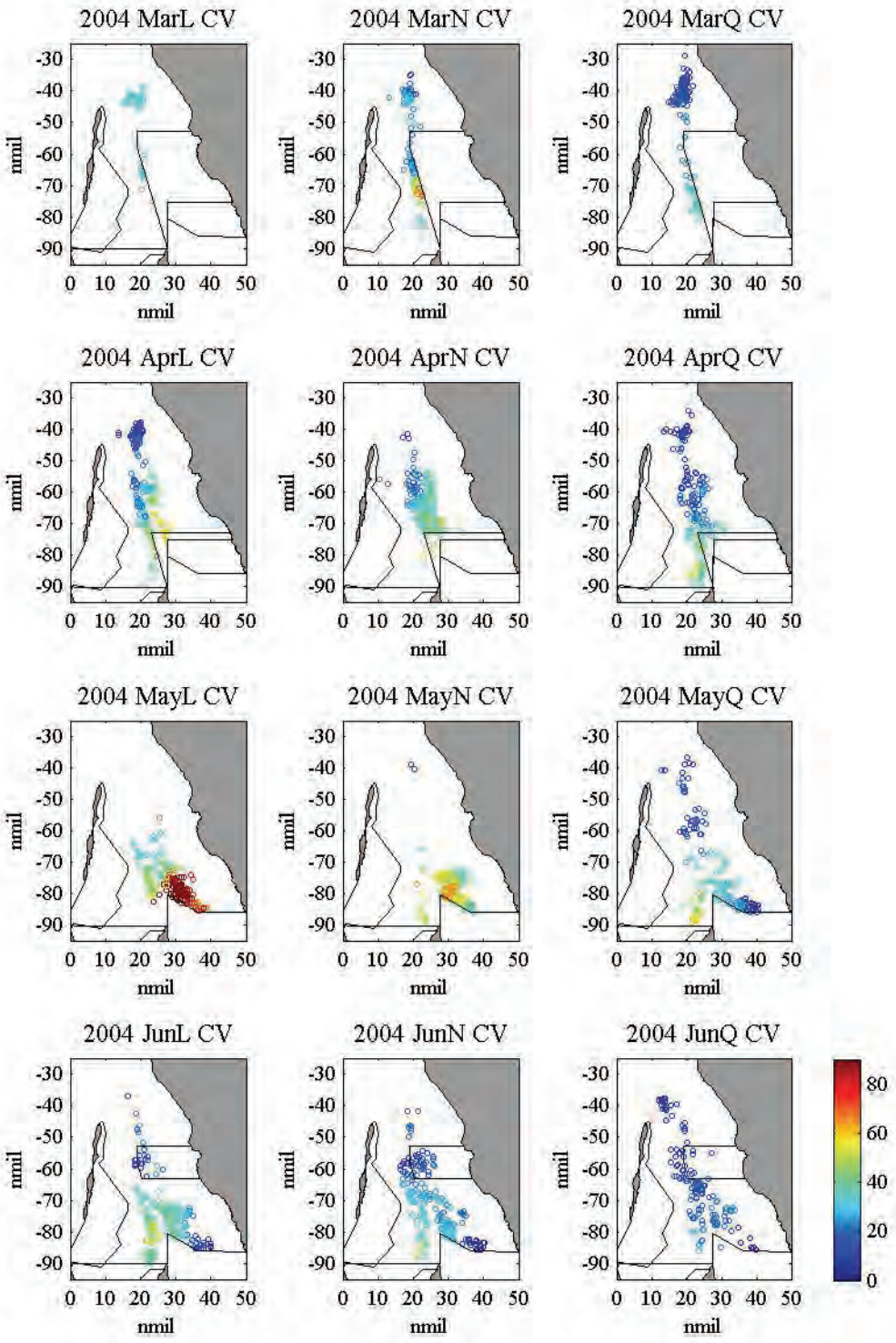
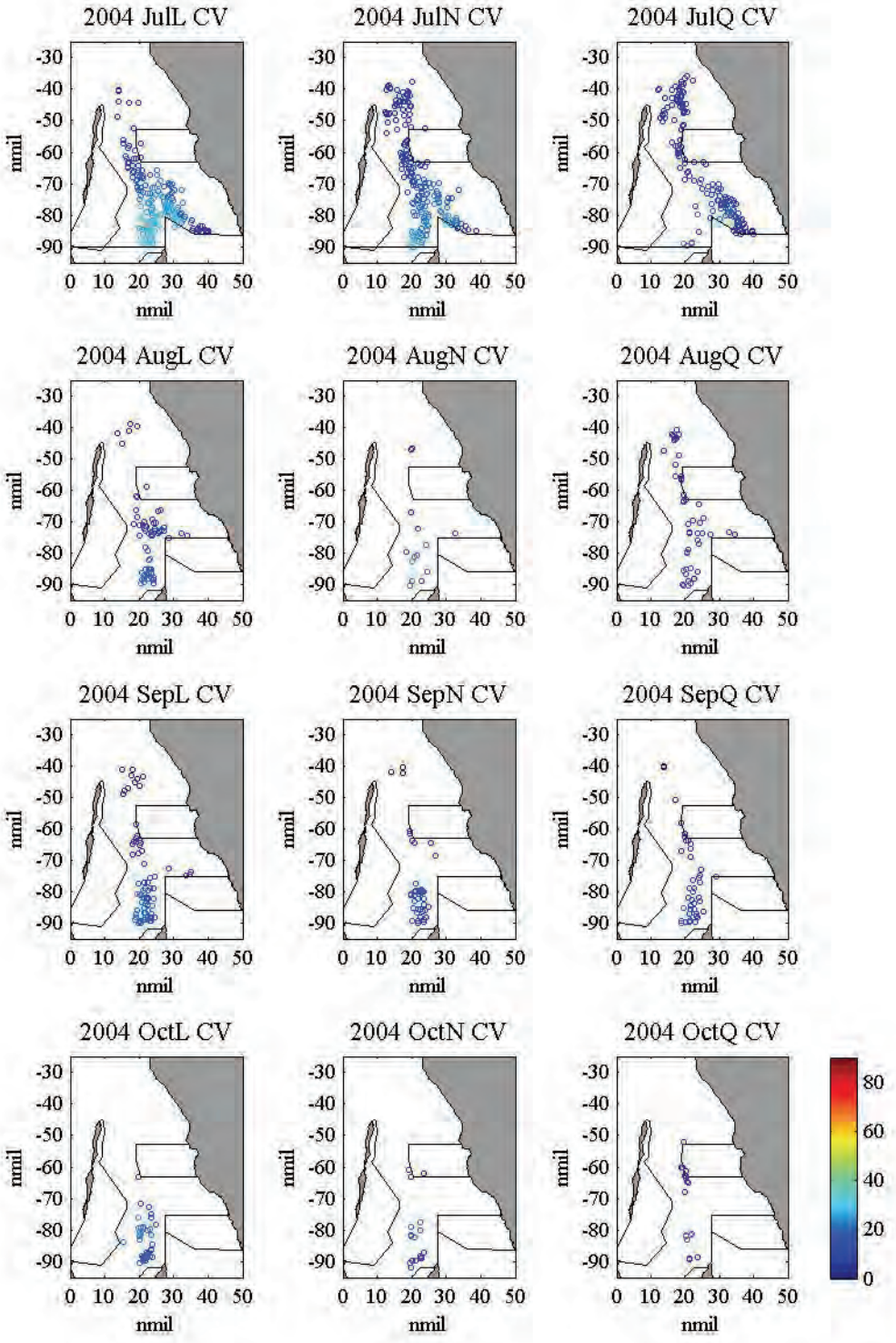


Figure J.7-2. Cross validation estimates of CPUE, July-October 2004.



J.8 Summary Statistics of CPUE Cross Validation Estimates 2004

Table J.8-1. Summary statistics for CPUE cross validation estimates, 2004.

Week	Mean	Med	Var	Skew	Min	Max	N	ME	MSE	MSDR	rho
MarL	35.89	34.05	45.25	3.04	29.24	63.98	51	-0.11	127.89	0.66	0.18
MarN	31.90	25.86	199.14	1.27	13.73	74.52	112	0.72	101.36	0.56	0.65
MarQ	18.35	15.83	67.31	1.27	6.90	41.16	140	0.17	31.57	0.25	0.66
AprL	27.23	24.88	203.09	0.46	7.16	61.23	200	0.16	90.40	0.29	0.68
AprN	33.24	34.65	79.03	0.01	14.09	60.99	202	0.01	124.93	0.45	0.30
AprQ	24.64	21.28	152.76	0.62	6.58	53.72	178	-0.59	73.46	0.46	0.68
MayL	68.64	74.93	527.65	-0.17	25.72	113.98	196	-0.94	365.30	1.37	0.61
MayN	47.69	46.46	92.85	-0.09	14.71	67.73	203	1.38	140.08	0.58	0.36
MayQ	25.49	22.66	171.57	0.93	6.21	62.74	165	-0.17	134.85	0.96	0.55
JunL	32.40	35.41	122.87	-0.57	5.06	55.74	198	0.36	96.96	0.37	0.53
JunN	23.12	23.93	89.71	0.05	2.30	52.23	195	-0.08	55.70	0.30	0.61
JunQ	16.37	17.81	33.80	-0.29	2.28	28.60	143	-0.44	33.31	0.33	0.48
JulL	21.47	22.04	50.84	-0.39	5.00	33.38	197	0.31	38.07	0.24	0.55
JulN	17.33	17.66	36.20	0.05	5.98	30.59	202	0.19	18.76	0.17	0.64
JulQ	11.36	9.96	20.40	0.56	4.12	22.95	166	-0.10	17.28	0.31	0.58
AugL	15.20	14.37	15.86	-0.30	5.14	22.56	69	0.40	24.95	0.30	0.34
AugN	17.67	16.96	17.34	0.29	10.86	25.95	14	-0.79	77.03	0.97	0.00
AugQ	12.67	12.59	9.81	0.19	6.93	18.36	48	-0.43	29.65	0.97	0.19
SepL	14.21	15.86	40.90	-0.40	3.01	22.34	74	0.39	13.81	0.27	0.76
SepN	14.47	16.12	19.16	-1.40	2.70	19.67	45	1.01	18.64	0.41	0.47
SepQ	12.52	12.74	16.47	-0.39	2.21	20.01	43	-0.59	32.95	1.19	0.16
OctL	18.93	18.15	11.08	-0.07	10.42	24.51	42	0.30	41.57	0.54	0.13
OctN	15.47	14.49	17.81	0.05	7.90	22.31	19	1.61	26.56	0.35	0.32
OctQ	12.36	11.47	25.57	0.69	6.49	21.89	24	-0.63	36.44	0.64	0.34

APPENDIX K: CPUE JACKNIFE AND GRID ESTIMATES

K.1 Jacknife and grid estimates, April 2004

Figure K.1-1. Jacknife and grid estimates of CPUE, April 2004.

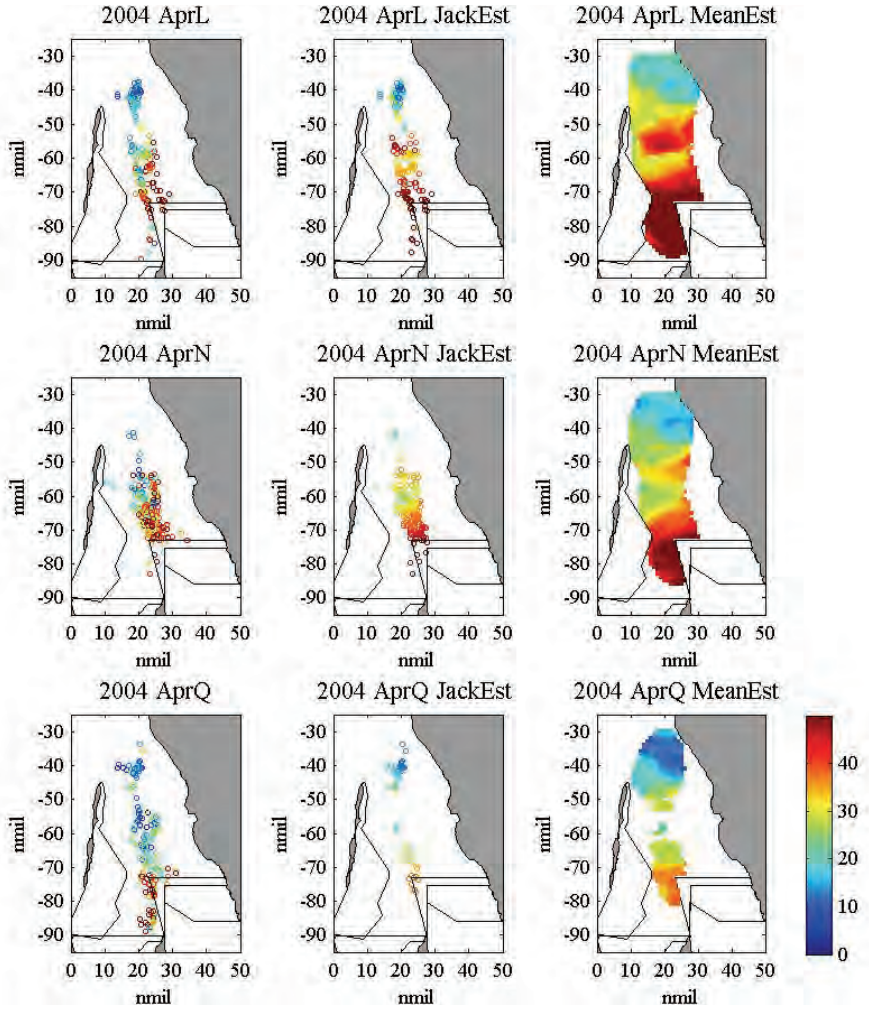
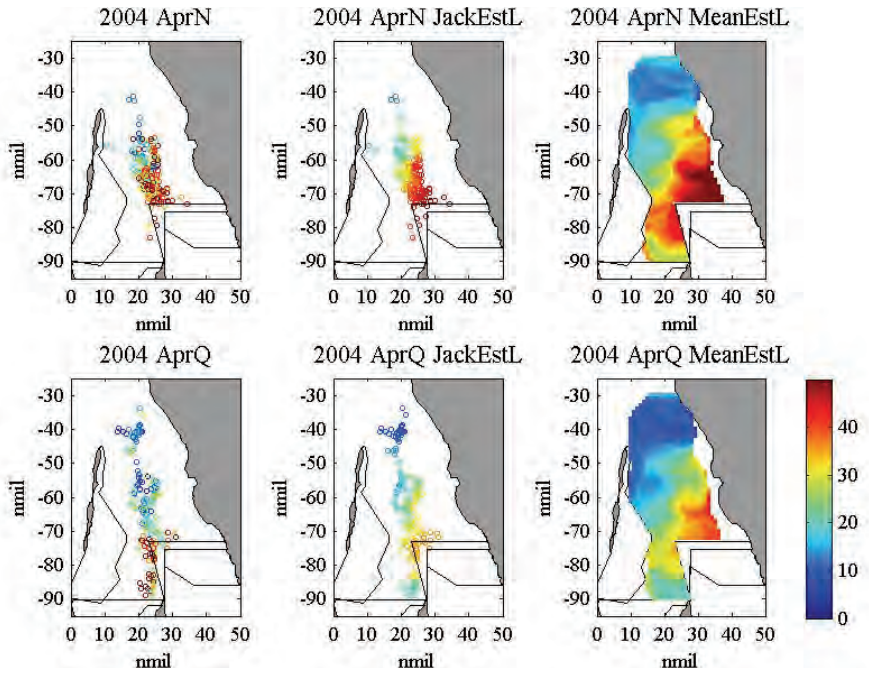


Figure K.1-2. Jacknife and grid estimates of CPUE using additional L week, April 2004.



K.2 Jacknife and grid estimates, June 2004

Figure K.2-1. Jacknife and grid estimates of CPUE, June 2004.

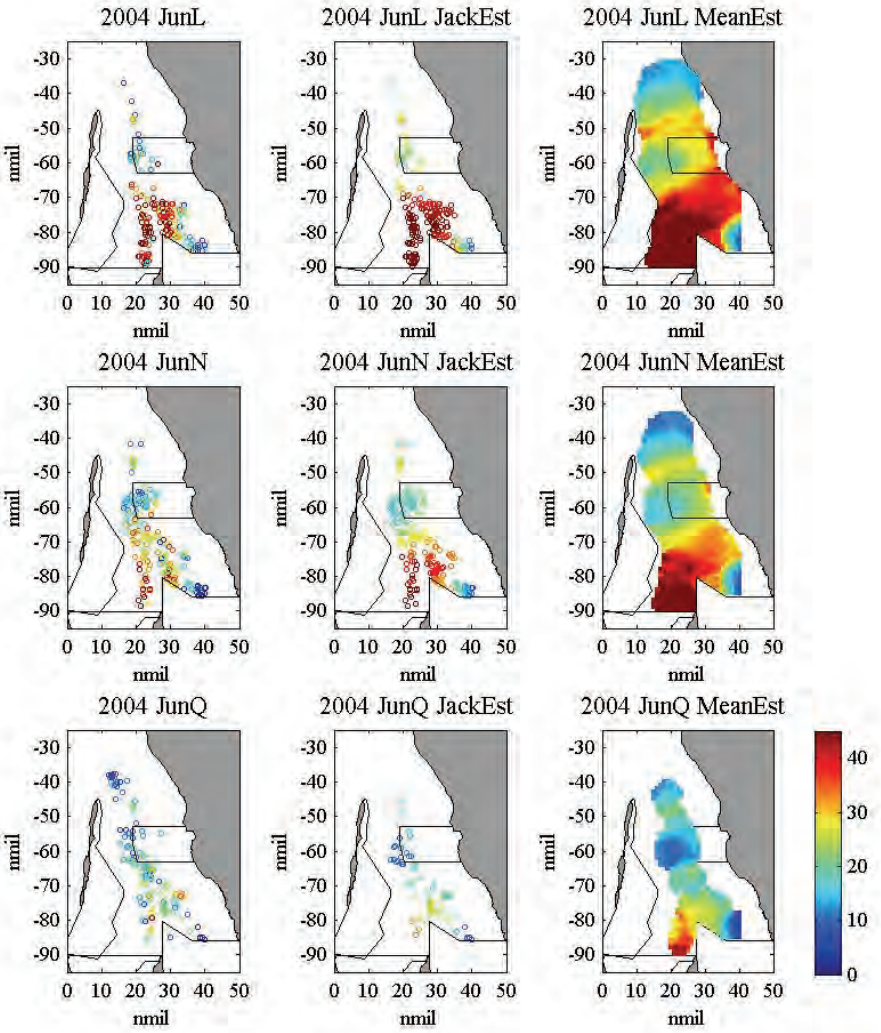
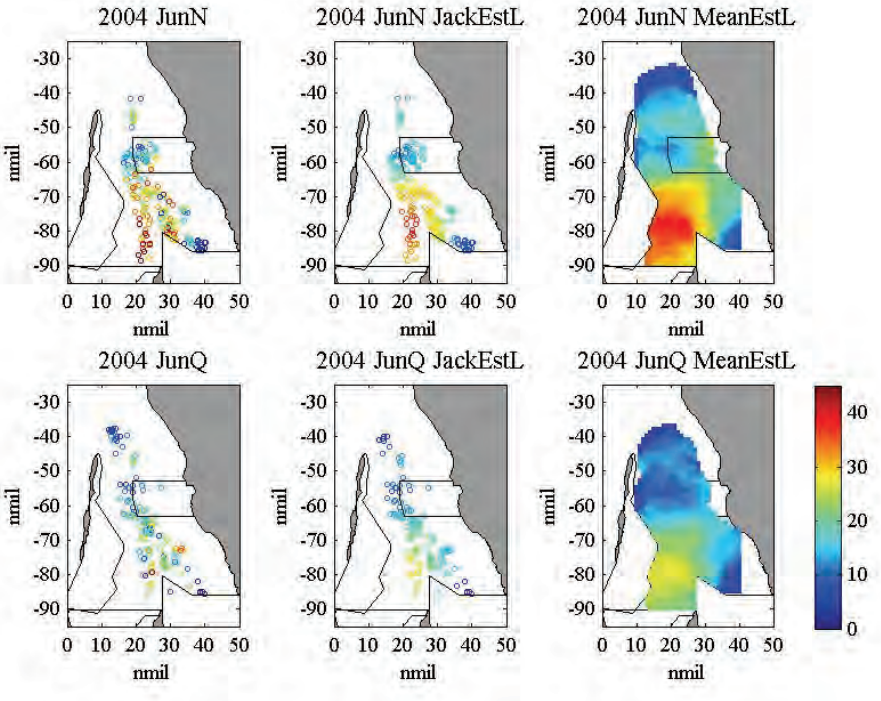


Figure K.2-2. Jacknife and grid estimates of CPUE using additional L week, June 2004.



K.3 Jacknife and grid estimates, August 2004

Figure K.3-1. Jacknife and grid estimates of CPUE, August 2004.

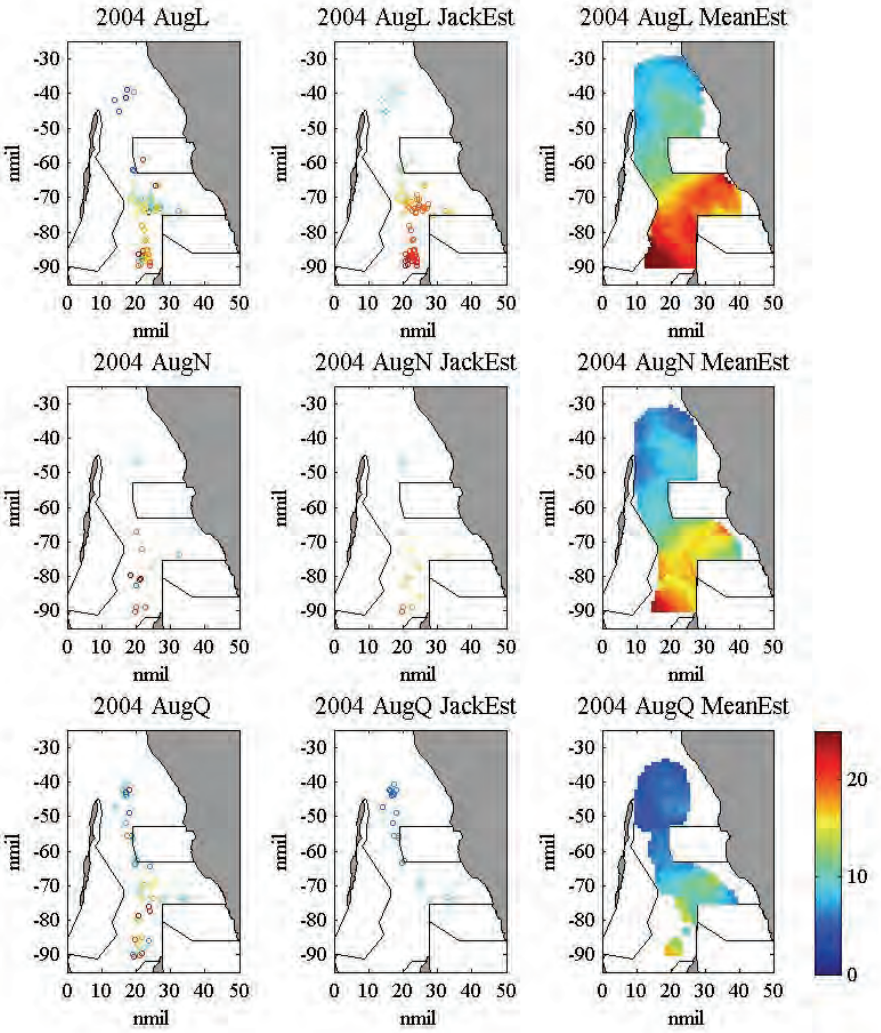
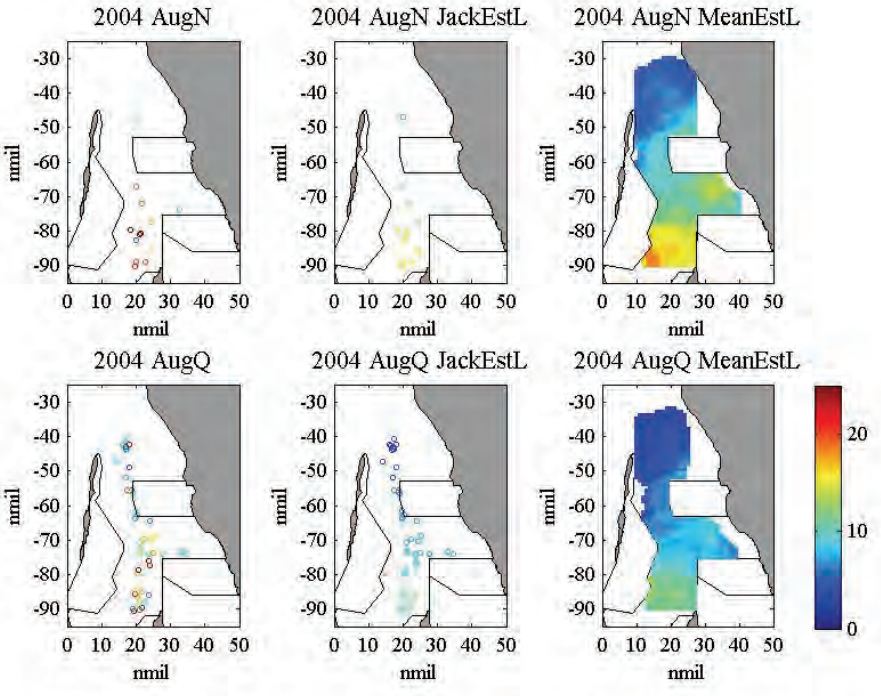


Figure K.3-2. Jacknife and grid estimates of CPUE using additional L week, August 2004.



K.4 Jacknife and grid estimates, September 2004

Figure K.4-1. Jacknife and grid estimates of CPUE, September 2004.

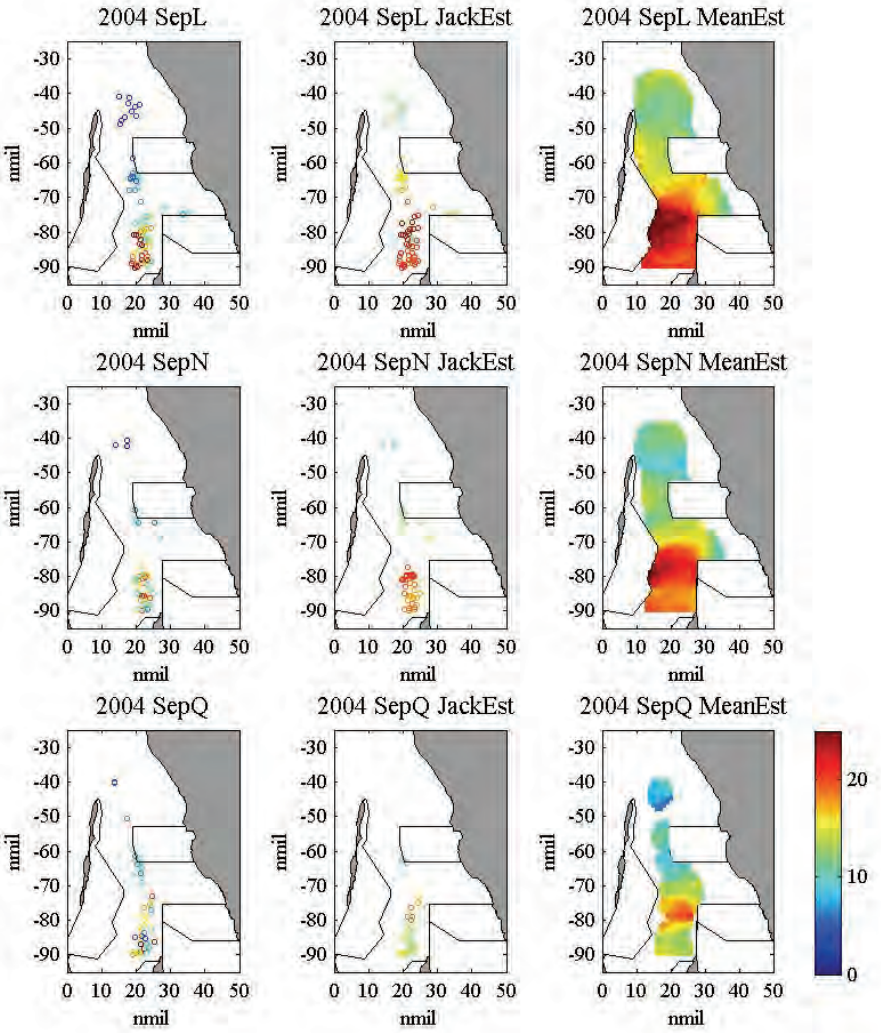
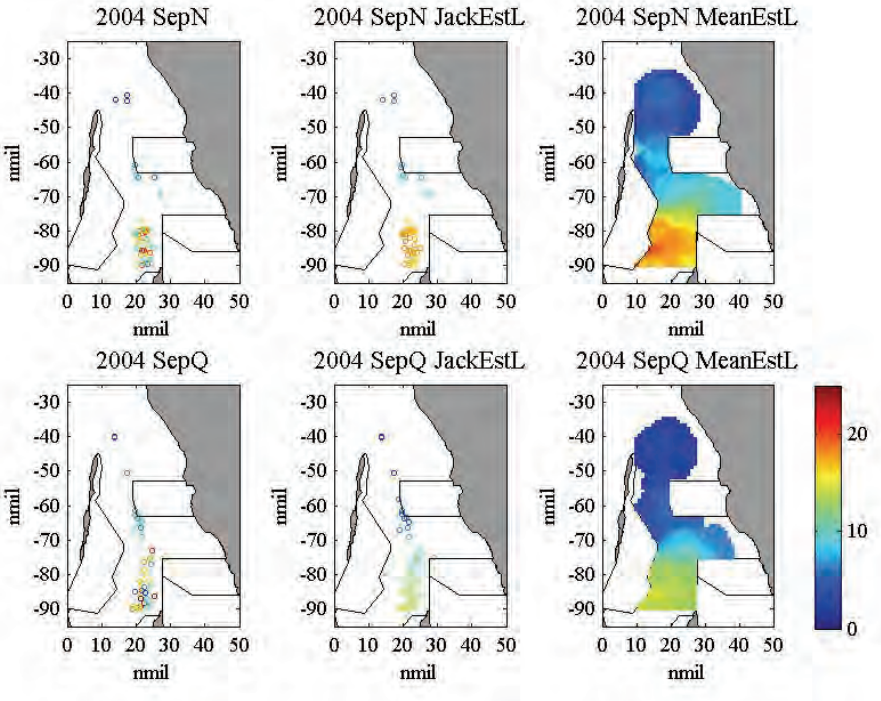


Figure K.4-2. Jacknife and grid estimates of CPUE using additional L week, October 2004.



K.5 Jacknife and grid estimates, October 2004

Figure K.5-1. Jacknife and grid estimates of CPUE, October 2004.

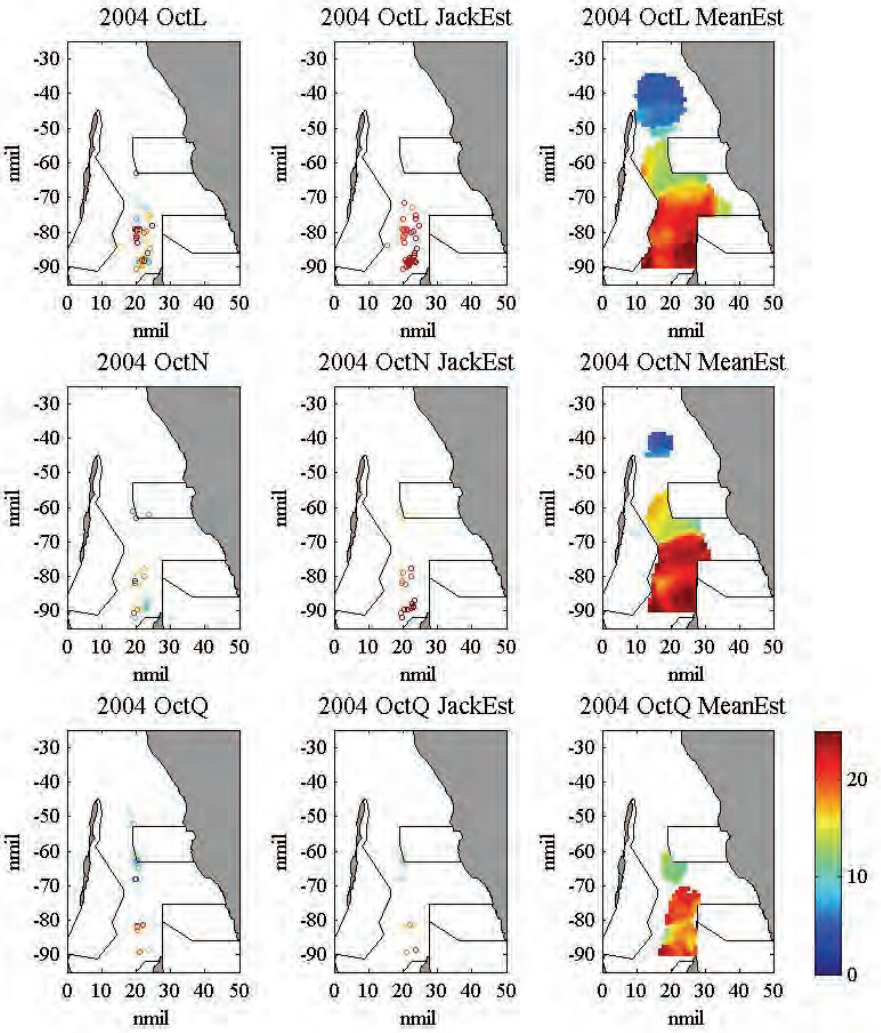
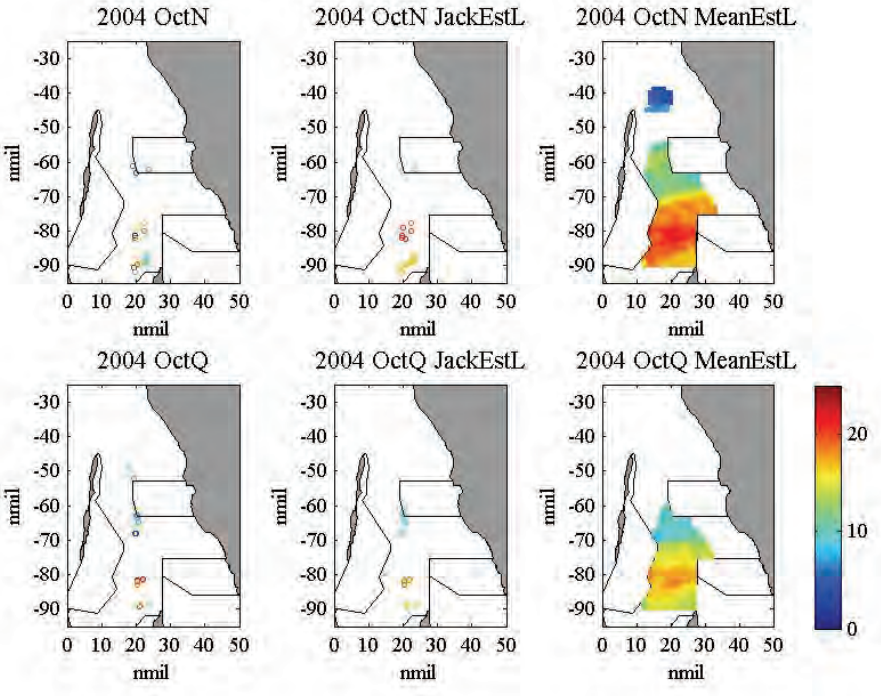


Figure K.5-2. Jacknife and grid estimates of CPUE using additional L week, October 2004.



K.6 Summary statistics of CPUE jacknife estimates

Table K.6-1. Summary statistics for CPUE jacknife estimates using previous month, 2004.

Week	Mean	Med	Var	Skew	Min	Max	N	ME	MSE	MSDR	rho
AprL	33.23	33.00	184.47	0.12	8.88	58.79	199	6.24	206.39	0.34	0.65
AprN	36.21	35.72	47.74	0.40	21.15	52.37	178	3.71	161.92	0.30	0.41
AprQ	21.67	19.17	84.90	0.57	10.99	39.58	56	0.35	73.70	0.21	0.78
MayL	45.99	47.79	102.91	-0.79	17.08	61.78	161	-21.75	1217.48	2.04	0.46
MayN	42.42	43.03	28.22	-2.75	12.82	52.42	135	-5.11	212.12	0.42	0.34
MayQ	22.65	20.87	116.80	0.30	5.40	41.06	63	-10.54	294.75	1.03	0.82
JunL	40.62	42.20	145.00	-0.41	10.19	64.37	198	8.57	204.74	0.42	0.64
JunN	27.81	27.09	105.67	0.32	8.18	60.31	195	4.61	101.09	0.29	0.68
JunQ	17.95	18.76	51.85	0.30	7.10	35.76	61	-1.08	67.89	0.33	0.47
JulL	24.13	25.70	56.02	-1.05	4.15	36.46	197	2.97	50.09	0.17	0.72
JulN	19.05	19.26	28.04	-0.02	4.14	31.27	198	1.88	27.44	0.13	0.73
JulQ	11.77	13.01	14.77	-0.80	1.15	18.54	55	-3.14	30.38	0.27	0.73
AugL	18.27	19.15	12.78	-1.10	8.73	24.21	69	3.48	35.85	0.23	0.61
AugN	14.86	15.52	12.57	-0.39	9.21	19.93	12	-1.41	20.77	0.17	0.66
AugQ	6.94	6.35	3.47	1.19	4.70	11.14	21	-3.72	35.15	0.56	0.25
SepL	18.15	19.62	17.30	-0.23	11.55	25.61	74	4.33	49.23	0.46	0.69
SepN	17.87	18.82	9.23	-1.01	10.57	21.69	45	4.40	40.08	0.47	0.61
SepQ	14.30	13.90	7.18	-0.15	8.21	18.96	27	-0.93	28.69	0.50	0.25
OctL	21.99	22.63	5.16	-1.68	13.01	24.57	42	3.36	62.09	0.42	0.04
OctN	21.80	22.82	16.22	-0.38	14.47	27.18	19	7.94	112.34	0.77	0.01
OctQ	14.94	14.14	12.33	0.42	10.88	20.99	11	1.49	33.84	0.29	0.64

Table K.6-2. Summary statistics for CPUE jacknife estimates using L week, 2004.

Week	Mean	Med	Var	Skew	Min	Max	N	ME	MSE	MSDR	rho
AprQ	23.62	24.46	79.40	-0.36	8.02	37.60	178	-1.61	162.79	0.53	0.55
MayQ	38.96	44.14	215.03	-0.95	5.40	58.52	159	13.10	537.76	2.01	0.30
JunQ	16.72	16.08	32.89	-0.08	4.45	28.62	131	-0.87	47.78	0.28	0.52
JulQ	9.39	9.03	9.49	0.46	3.82	17.91	164	-2.18	28.22	0.29	0.67
AugQ	7.96	7.72	6.95	0.12	3.86	12.06	48	-5.14	56.40	0.94	0.41
SepQ	10.76	12.97	13.02	-1.03	2.15	13.98	43	-2.35	32.00	0.66	0.52
OctQ	13.21	12.44	15.53	0.27	8.68	18.58	18	-0.25	16.91	0.18	0.89

K.7 Summary statistics of CPUE grid estimates

Table K.7-1. Summary statistics for CPUE grid estimates using previous month.

Week	Mean	Median	Variance	Skewness	Min	Max	Count
AprL	35.198	33.536	143.796	0.196	13.689	59.763	6951
AprN	29.900	27.471	117.105	0.490	10.068	56.816	5281
AprQ	20.433	18.767	73.398	0.657	8.138	40.568	1279
MayL	34.479	29.286	240.219	0.489	8.648	80.103	9392
MayN	28.923	24.360	185.075	0.477	6.901	71.007	7437
MayQ	19.657	17.021	97.211	0.498	4.979	45.232	2330
JunL	32.440	29.404	164.868	0.506	8.946	68.377	9400
JunN	27.735	25.490	119.863	0.606	7.922	63.533	6983
JunQ	18.289	17.513	54.218	0.611	5.116	43.575	1453
JulL	21.257	21.185	55.679	0.109	3.383	43.893	8450
JulN	17.082	17.818	31.959	-0.039	2.353	35.292	6936
JulQ	11.841	12.439	14.437	-0.093	1.164	23.071	1820
AugL	14.250	11.776	29.166	0.391	6.046	29.892	7920
AugN	11.373	9.517	17.626	0.476	4.877	24.221	5650
AugQ	7.252	6.338	6.417	0.991	3.824	17.036	2105
SepL	16.423	15.143	19.271	0.597	8.645	30.304	6556
SepN	15.278	14.456	17.765	0.324	7.015	26.195	5034
SepQ	13.582	13.490	10.763	-0.011	5.986	22.042	1039
OctL	16.182	17.515	43.394	-0.467	3.543	29.610	4728
OctN	20.158	21.570	21.133	-1.033	4.583	28.606	2424
OctQ	17.853	18.937	11.961	-0.483	10.883	24.163	403

Table K.7-2. Summary statistics for CPUE grid estimates using additional L week.

Week	Mean	Median	Variance	Skewness	Min	Max	Count
AprQ	21.009	20.509	83.571	0.355	6.246	45.462	7677
MayQ	33.805	37.959	214.504	-0.365	4.979	60.924	5522
JunQ	16.152	14.471	35.335	0.366	4.401	28.551	7163
JulQ	10.581	10.632	14.516	0.339	3.805	18.276	6370
AugQ	7.290	7.412	7.258	0.140	1.820	14.051	5131
SepQ	7.068	6.233	18.266	0.426	1.592	15.168	5483
OctQ	14.661	15.315	6.936	-0.957	7.296	19.317	2680