

**DETERMINATION OF GEO-STATISTICAL TECHNIQUE FOR
OPTIMISATION OF SPATIAL OCEANOGRAPHIC DATA**

AZIZI BIN ALI

**Thesis Submitted in Fulfilment of the Requirement for the
Degree of Doctor of Philosophy in the
Institute of Oceanography and Environment
Universiti Malaysia Terengganu**

Mei 2015

DEDICATION

This thesis is dedicated to my beloved wife Rosmawati Yusof who share a sad time and laughs time in journey, to my sons Adzraf Syaifullah, Anas Syariful and Ammar Syakirin and my daughter Amni Syahirah who has become a source of motivation and ispirasi. Always give joy and peace when are facing difficult times in the course of seeking knowledge. Dedication also for my mother Minah @ Tengah Harun and in law family, Yusof Ali and Ramlah who always pray in this journey.

Finally, this thesis is dedicated to all those who believe in the richness of learning.

Abstract of thesis presented to the Senate of Universiti Malaysia Terengganu in fulfillment of the requirement for the degree of Doctor of Philosophy

**DETERMINATION OF GEO-STATISTICAL TECHNIQUE FOR
OPTIMISATION OF SPATIAL OCEANOGRAPHIC DATA**

AZIZI BIN ALI

May 2015

Main supervisor : **Assoc. Prof. Aidy @ Mohamed Shawal Bin M
Muslim, Ph.D.**

Members : **Mohd Fadzil Bin Mohd Akhir, Ph.D**
: **Prof. Mohd Lokman Husain, Ph.D**

Institute : **Institute of Oceanography and Environment**

The main challenge in most of the marine environmental monitoring studies has always been the difficulties and high cost during data collection and analysis process. Therefore, researchers have to carefully plan such as in determine the suitable sampling scheme and total number of sampling station in order to produce the good results. To counter this inadequacy, researchers usually have to use prediction and estimation data within the sampling location in order to produce better data representation of the study area. Therefore, important for researchers to determine the most suitable error estimation techniques used especially in spatial monitoring where interpolation dataset will be used due to the limited field sampling data.

This thesis presents the implementation of geostatistical techniques as a tool in estimation and mapping the spatial distribution of the oceanography parameters. Geostatistical is spatial interpolation technique employed to predict the measurement of parameters at unsampled areas by using the values measured at other locations with existence in situ data. There are three main physical parameters have been selected in this study which include sea surface temperature (SST), sea surface salinity (SSS) and chlorophyll-a. The study area was divided into two main areas; first is in offshore of Terengganu water of South China Sea (SCS) water. Second study area is located at costal of Terengganu water. There were three main analyses in this study; firstly to analysed for investigates and established the suitable model using geostatitical technique and compared with classical estimation techniue. Secondly, to determining the most accurate sampling in term of error produce using the geostatistical and classical approaches. Thirdly, to determine the effective sampling pattern to mapping the spatial distribution of the ocean parameters in term of error produces. There are four sampling pattern will be analysis included systematic scheme, random scheme, stratified scheme and cluster scheme.

The result shows that the kriging estimation technique was the most suitable method which produced the lowest error for mapping the distribution of the scale space region, in comparison with the classical techniques (IDW and Spline). The results from the analysis of the model as demonstrated that the spherical model is the most suitable model for sea surface temperature and chlorophyll-a dataset, whereas the exponential model is much better for sea surface salinity data. The results also revealed that even though the inversed distance weight and spline methods have the advantage of simplicity and easy to process, but these methods are less accurate.

Overall the percentage of root mean square error (RMSE) difference for SST between the kriging technique with spherical model in comparison with IDW techniques at power of three and spline technique is as follows: Kriging was compared to IDW is 51.2% and Kriging was compared to spline is 58.7%

In determining sampling density optimisation techniques were utilized, the sampling density (station) was reduced in stages to determine the most stable error estimation. Results show that within the coastal waters of Terengganu, RMSE generated using kriging technique produced the lowest density with no significant error at a density of 90 stations (1.3km station distributed) when compared to the initial 314 sampling stations (0.382km station distributed). Therefore it was determined that 90 stations is the optimal density in this case.

In investigate the most accurate sampling pattern, results shows that systematic sampling scheme produced less error compared to other method. For instance, in sea surface temperature parameter, at such systematic scheme was proven to be the best scheme in 1.5km x 1km grid interval at 103 sampling stations density. Differences in the percentage of sampling error compared to another scheme at density is 103 are follows; (percentage error of 9.3% for stratified scheme, 11% for cluster scheme and 11% random scheme). The percentage of difference between sampling grids in the systematic scheme also shows grid 1.5km x1km at 120km² study areas found less error compared to another grid. The error difference between 1.5km x 1km grids with is 2km x 2km about 19.5% and 1.5km x 1km between the 2km x 1km is 15.9%. Therefore, the systematic sampling scheme was found to be the most suitable sampling pattern in sea surface temperature at study area.

Abstract of thesis presented to the Senate of Universiti Malaysia Terengganu in fulfillment of the requirement for the degree of Doctor of Philosophy

**PENENTUAN TEKNIK GEO-STATISTIK UNTUK MENGOPTIMUNKAN
SPATIAL DATA OSEANOGRAFI**

AZIZI BIN ALI

Mei 2015

**Penyelia Utama : Assoc. Prof. Aidy @ Mohamed Shawal Bin M
Muslim, Ph.D.**

Penyelia Bersama : Mohd Fadzil Bin Mohd Akhir, Ph.D
: Prof. Mohd Lokman Husain, Ph.D

Institut : Institut Oseanografi dan Sekitaran

Cabaran utama dalam kebanyakan kajian pemantauan alam sekitar marin sentiasa menghadapi kesukaran dan kos yang tinggi dalam pengumpulan data dan proses analisis. Oleh itu, penyelidik perlu berhati-hati merancang seperti dalam menentukan skim pensampelan yang sesuai dan jumlah bilangan stesen persampelan untuk menghasilkan keputusan yang baik. Untuk mengatasi kekurangan ini, penyelidik biasanya perlu menggunakan ramalan dan anggaran data dalam lokasi persampelan bagi melahirkan perwakilan data yang lebih baik daripada kawasan kajian. Oleh itu, penting bagi penyelidik untuk menentukan teknik anggaran ralat paling sesuai digunakan terutama dalam memantau ruang di mana interpolasi dataset akan digunakan kerana terhad data bidang persampelan.

Tesis ini membentangkan pelaksanaan teknik Geostatistik sebagai alat dalam membuat anggaran dan pemetaan taburan parameter oseanografi. Geostatistik adalah teknik interpolasi ruang digunakan untuk meramalkan pengukuran parameter di kawasan unsampled dengan menggunakan nilai-nilai yang diukur di lokasi-lokasi lain dengan kewujudan dalam data situ. Terdapat tiga parameter fizikal utama telah dipilih dalam kajian ini termasuk suhu permukaan laut (SST), kemasinan permukaan laut (SSS) dan klorofil-a. Kawasan kajian telah dibahagikan kepada dua kawasan utama; pertama adalah di luar pesisir air Terengganu Laut China Selatan (SCS) dan kawasan kajian kedua terletak di persisiran pantai perairan Terengganu. Terdapat tiga analisis utama dalam kajian ini; pertama untuk dianalisis untuk menyiasat dan ditubuhkan model yang sesuai menggunakan teknik geostatistical dan dibandingkan dengan anggaran klasik technique. Kedua, untuk menentukan persampelan yang paling tepat dari segi hasil kesilapan menggunakan pendekatan geostatistik dan teknik klasik. Ketiga, untuk menentukan corak pensampelan berkesan untuk pemetaan taburan parameter laut dari segi kesilapan menghasilkan. Terdapat empat pola persampelan akan analisis termasuk skim sistematik, skim rawak, skim berstrata dan skim kelompok.

Hasil kajian menunjukkan bahawa teknik geostatistik dikenali kriging interpolasi adalah kaedah yang paling sesuai yang menghasilkan ralat yang paling rendah untuk pemetaan pembahagian rantau ruang skala, dalam perbandingan dengan teknik klasik (IDW dan spline). Hasil daripada analisis model sebagai menunjukkan model sfera adalah model yang paling sesuai untuk suhu permukaan laut dan klorofil-a dataset, manakala model eksponen adalah jauh lebih baik untuk data kemasinan permukaan laut. Kajian ini juga mendapati bahawa walaupun kaedah IDW dan Spline

mempunyai kelebihan kesederhanaan dan mudah untuk proses, tetapi kaedah ini adalah kurang tepat. Keseluruhan peratusan perbezaan RMSE untuk SST antara teknik 'kriging' dengan model sfera berbanding dengan teknik IDW pada kuasa tiga dan teknik Splin adalah seperti berikut: 'kriging' dibandingkan dengan IDW adalah 51.2% dan 'kriging' dibandingkan dengan Splin adalah 58.7%

Dalam menentukan teknik pengoptimuman ketumpatan sampel telah digunakan, ketumpatan pensampelan (stesen) telah dikurangkan secara berperingkat-peringkat untuk menentukan anggaran ralat yang paling stabil. Hasil kajian menunjukkan bahawa dalam kawasan perairan Terengganu, RMSE dijana menggunakan teknik 'kriging' dihasilkan ketumpatan paling rendah tanpa kesilapan penting di ketumpatan 90 stesen (1.3 stesen diedarkan) berbanding awal 314 stesen persampelan (stesen 0.382km diedarkan). Oleh itu ia telah ditentukan bahawa 90 stesen adalah ketumpatan optimum dalam kes ini. Setelah menentukan ketumpatan sampel optimum dan corak (skim sistematik) analisis lanjut dijalankan untuk menentukan selang saiz grid yang paling sesuai untuk perancangan bidang sebenar.

Dalam menentukan corak persampelan yang paling tepat, keputusan menunjukkan bahawa skim pensampelan sistematik dihasilkan ralat kurang berbanding dengan kaedah lain. Sebagai contoh, dalam parameter suhu permukaan laut, di rancangan sistematik itu telah terbukti skim yang terbaik di 1.5km x 1km selang grid di 103 stesen pensampelan kepadatan. Perbezaan peratusan ralat persampelan berbanding dengan skim lain pada kepadatan adalah 103 adalah seperti berikut; (ralat peratusan sebanyak 9.3% bagi skim berstrata, 11% bagi skim kelompok dan 11% skim rawak). Peratus perbezaan antara grid persampelan dalam skim sistematik juga menunjukkan

grid 1.5 x1km di 120km² kawasan kajian mendapati ralat kurang berbanding grid lain. Perbezaan ralat antara 1.5 x 1km grid dengan adalah 2km x 2km tentang 19.5% dan 1.5 x 1km antara 2km x 1km adalah 15.9%. Oleh itu, skim pensampelan sistematik didapati corak persampelan yang paling sesuai dalam suhu permukaan laut di kawasan kajian.