

Estimação da abundância por idades, da
Pescada (*Merluccius merluccius*),
usando métodos geoestatísticos bayesianos e
análise de dados composicionais

Ernesto Jardim (IPIMAR)

<ernesto@ipimar.pt>

&

Paulo Ribeiro Jr (UFPR)



→ Resumo

- 1) Introdução: objectivos
- 2) Material: campanhas demersais IPIMAR e Pescada;
- 3) Métodos: decomposição do índice de abundância, métodos estatísticos, pressupostos;
- 4) Resultados: composições por idades, índice abundância anual, distribuição espacial de Pescada, índice abundância anual por idade;
- 5) Discussão:

→ Introdução

Motivação:

Necessidade de obter indicadores da abundância por idade dos recursos pesqueiros, para utilização na avaliação e posterior aconselhamento sobre exploração dos mesmos, considerando os vários aspectos que influenciam a abundância (espaço/tempo) e contabilizando “correctamente” a incerteza associada.

Produção de dados para ambientes de simulação como Avaliação de Estratégias de Gestão.

Alguns problemas técnicos relacionados com o tipo de dados recolhidos nas campanhas demersais (sobredispersão, esforço amostral reduzido).

→ Introdução

Objectivos:

Propôr metodologia estatística para estimar índices de abundância por idade, combinando a distribuição espacial do recurso e a sua estrutura populacional.

Propôr métodos de simulação para obter distribuição estatística completa dos índices de abundância.

Explorar soluções para questões técnicas relacionadas com sobredispersão e grande quantidade de observações nulas.

→ Material

Campanhas demersais:

Na costa continental desde 1979 com NI Noruega e NI Capricórnio.

Estimar índices de abundância e biomassa e distribuição espacial das espécies comerciais mais importantes. Obter parâmetros biológicos individuais de pescada, carapau, verdinho, areeiros e tamboris.

Desenho amostral aleatório estratificado, arrastos de 30 minutos. Rede NCT 180/96 com saco de 20mm de malhagem. Abertura vertical de 4.8m e horizontal de 15.6m.

Série Outono dirigida a estimar abundância de recrutas. Série Primavera dirigida a estimar abundância de adultos.

Dados: observações de número de pescadas de uma determinada idade capturadas por hora, data/hora, localização geográfica, profundidade, tempo de arrasto, salinidade e temperatura de fundo.

→ Material

Pescada:

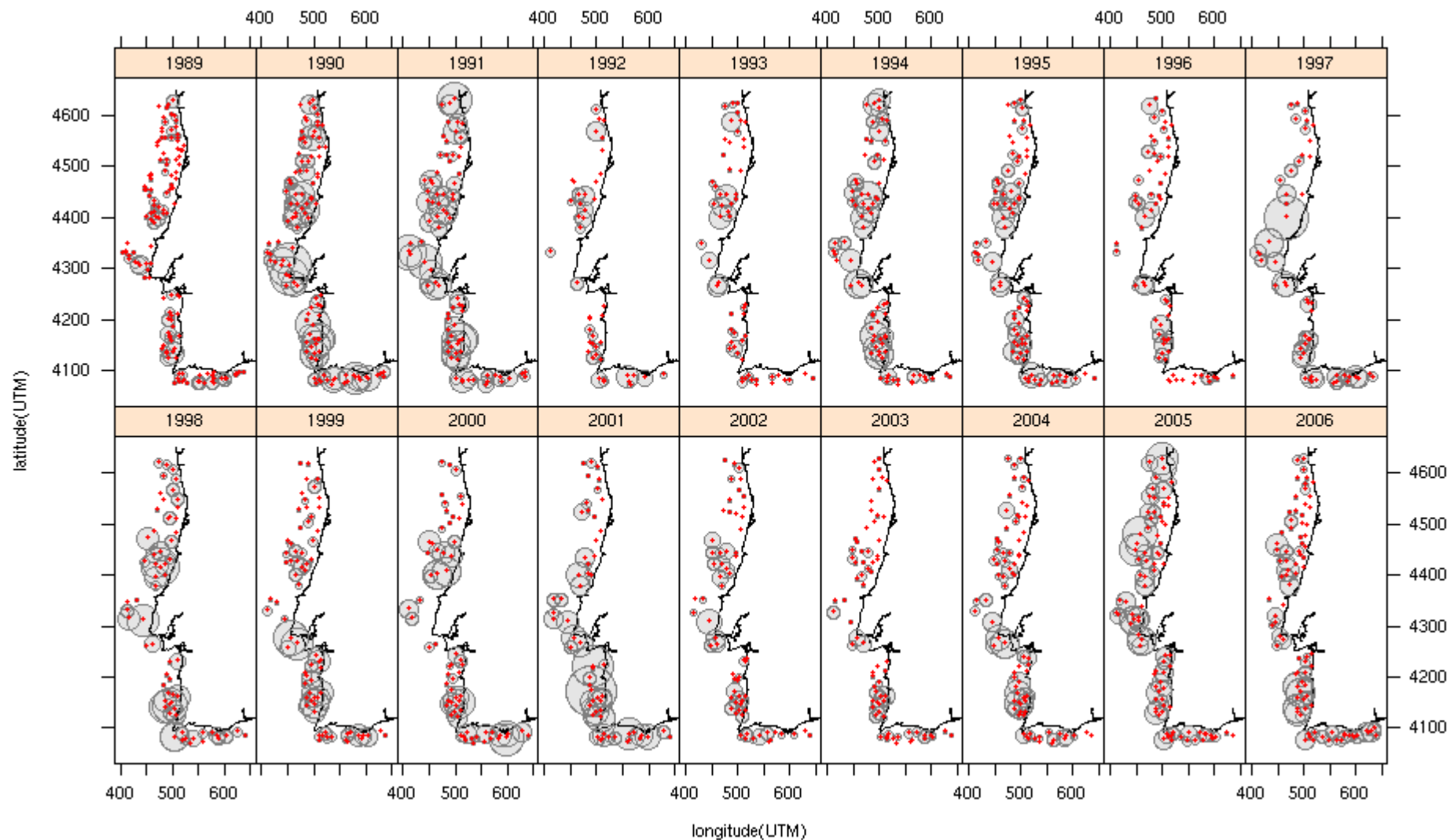
Existe em toda a costa europeia atlântica. Em Portugal desova entre Dezembro e Abril, recruta à pesca entre Setembro e Novembro.

Muito importante comercialmente. Capturada pelas frotas arrasto, polivalente e palangre. Existe em toda a costa continental e é desembarcada em quase todos os portos. Desembarques à volta de 2000 toneladas ano (40% arrasto, 50% polivalente, 10% palangre).

Unidade de gestão Ibérica fora dos limites biológicos de segurança, sujeita a plano de recuperação.

→ Material

Yearly maps with locations of hauls (+) and observed catches of Hake during the Autumn series of the Portuguese bottom trawl survey. The gray circles are proportional to the logarithm of the numbers of individuals caught per hour.



→ Métodos

Considerando:

$$i=1, \dots, n(\text{idade})$$

$$j=1, \dots, m(\text{ano})$$

I_{ij} = abundância por idade

P_{ij} = proporção de indivíduos por idade

$$Y_j = \sum_i I_{ij} = \text{abundância total}$$

$$I_{ij} = Y_j P_{ij}$$

Factorização permite utilizar métodos adequados ao tipo de dados.

Pressuposto de independência entre Y e P.

→ Métodos

P_i foi modelado usando “compositional data analysis” (Aitchison 1982, 2003), em particular a transformação “Additive log-ratio” que transforma um vector de proporções (composição) em dados gaussianos multivariados.

- 1) transformar dados
- 2) calcular parâmetros da gaussiana multivariada
- 3) reamostrar com bootstrap paramétrico 1000 vezes
- 4) transformar simulações para a escala original
- 5) calcular estatísticas de interesse, mediana, intervalos de confiança, etc.

→ Métodos

Y_i foi modelado usando geoestatística “Model Based Geostatistics” (Diggle et al. 1998; Diggle and Ribeiro 2007) de forma a considerar explicitamente a correlação espacial entre localizações, recorrendo a metodologia bayesiana para considerar a incerteza nos parâmetros.

$Z(x)$ = observações com ruído de um processo estocástico

x = localizações em R^2

$$Z(x) = S(x) + \varepsilon \quad \varepsilon = \text{Gau}(0, \tau^2)$$

$S(x)$ processo estocástico estacionário

$$E[S(x)] = \beta \quad \text{Var}[S(x)] = \sigma^2 \quad \rho(h) = \text{Corr}[S(x), S(x')]$$

$$Z(x) \sim \text{MVG}(\beta, \theta) \quad \theta = (\sigma^2, \varphi, \tau^2)$$

→ Métodos

Análise geoestatística:

1) GLM com erros binomais negativos e função de ligação log para calibrar observações e tratar problema de sobre dispersão

$$Z(x) = f(Y(x))$$

2) função de correlação exponencial

3) definir prioris para θ (iguais todos os anos)

4) obter posterioris

5) previsão (kriging)

6) simulação condicional

7) estimar índice anual $Y_i = \exp(\beta)$

8) estimar distribuição espacial $Z(x_0)$

→ Métodos

Considerando:

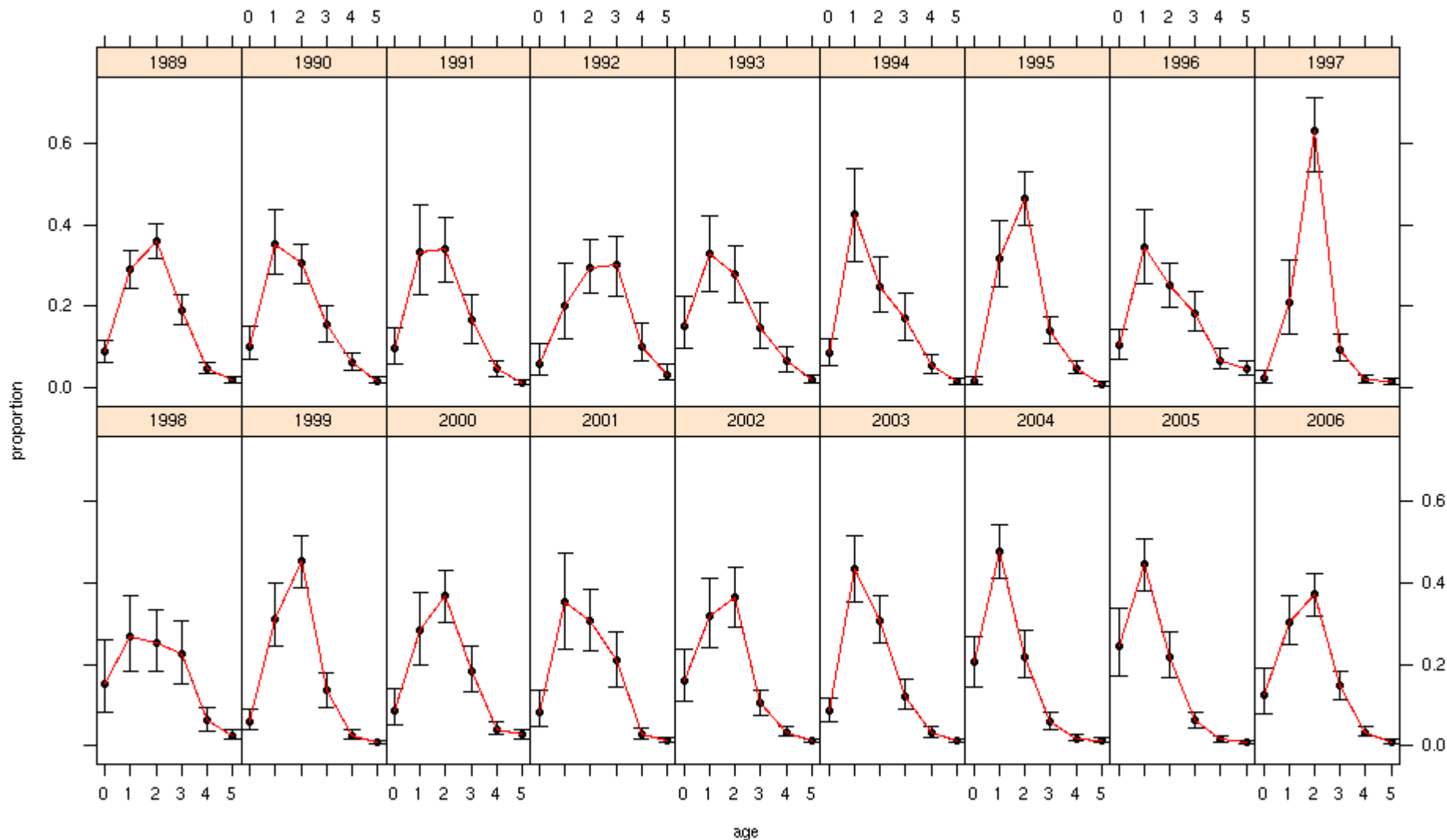
$$s=1, \dots, S(\text{simulação})$$

$$I_{ijs} = Y_{is} P_{ijs}$$

Simulações permitem obter a distribuição do índice de abundância e abrem perspectivas novas de análise dos resultados.

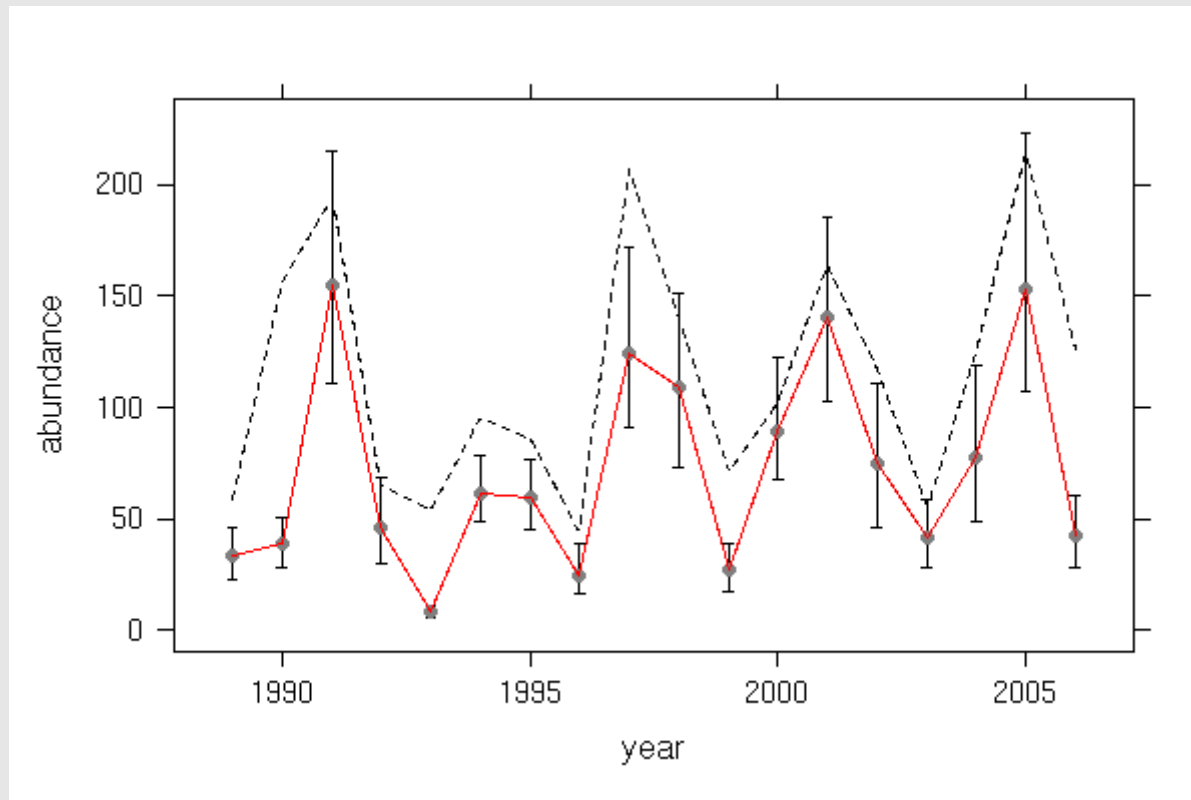
→ Resultados (P_i)

Age compositions empirical distribution obtained by parametric bootstrap. The full circle represents the median proportion and the gray lines represent the confidence interval computed by the 0.025 and 0.975 percentiles.



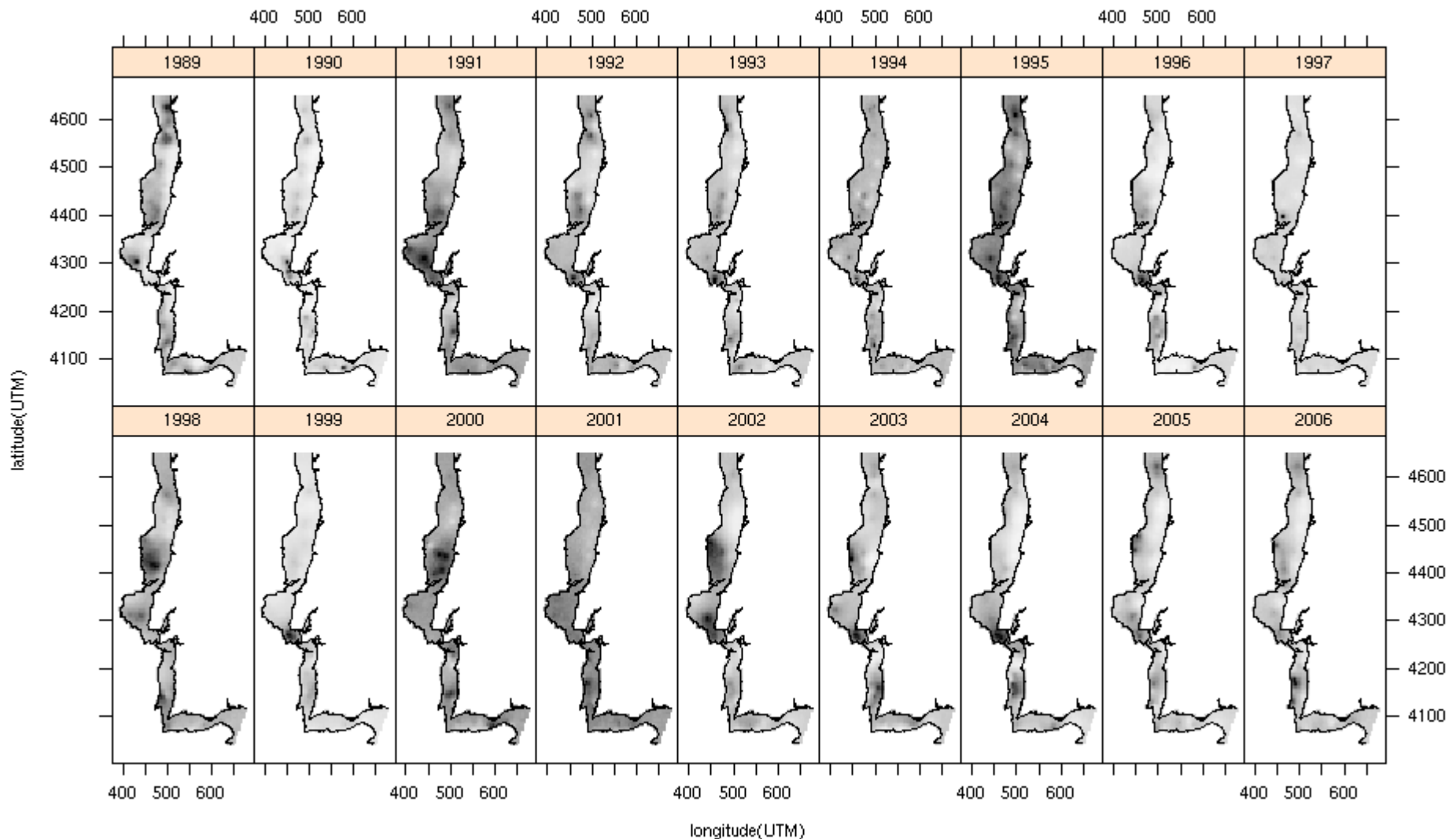
→ Resultados (Y_i)

Yearly abundance estimates by design statistics (dashed line) and geostatistics (full line). The black circle represents the median abundance and the gray lines represent the confidence interval computed by the 0.025 and 0.975 percentiles.



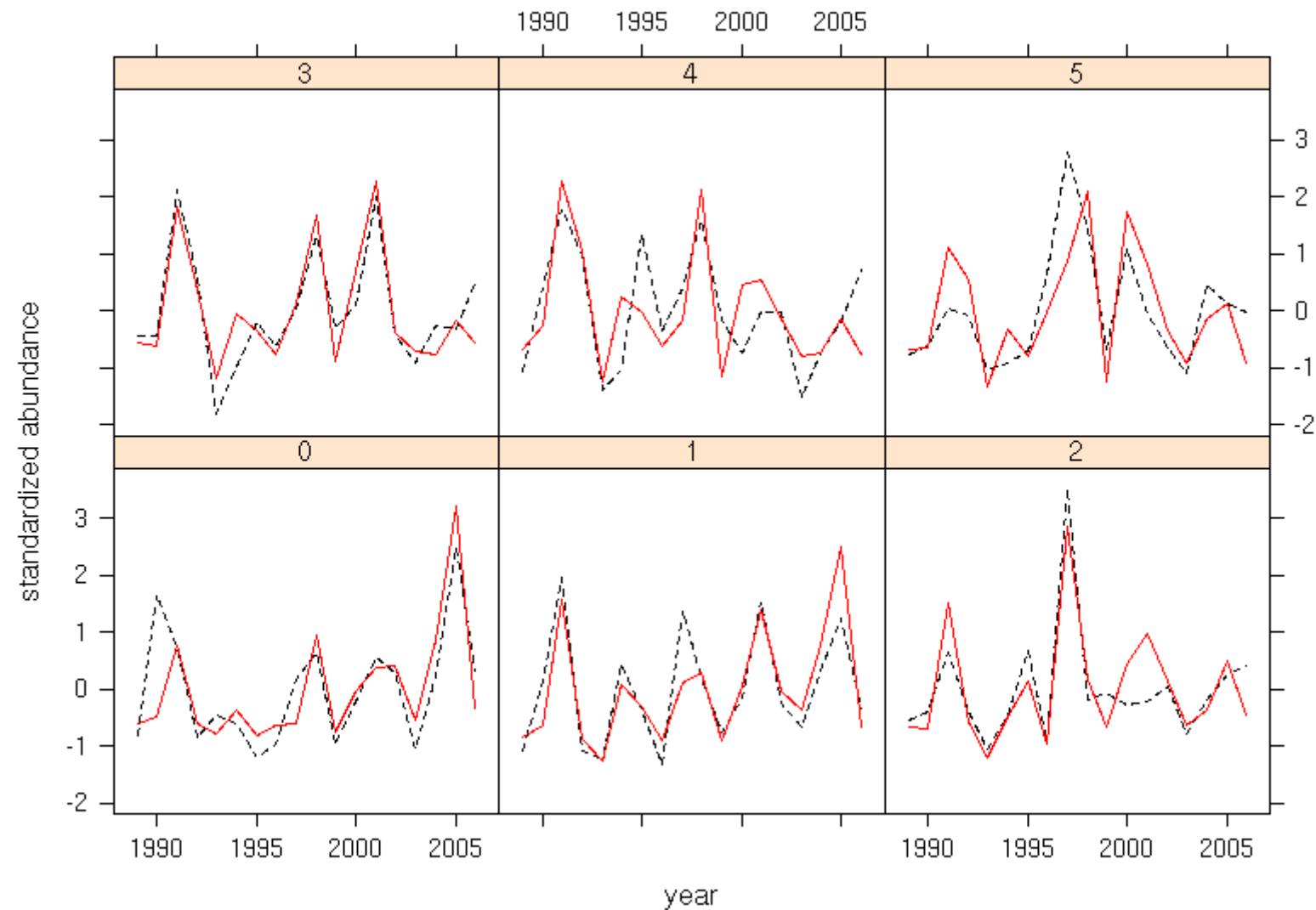
→ Resultados ($Z(x_0)$)

Spatial distribution of abundance by year standardized to the second fortnight of October. The gray degrees are proportional to the number of individuals caught per unit effort, rescaled to the maximum estimate within each year. The black = 1; white = 0.



→ Resultados (I_{ij})

Abundance at age and year standardized to have mean 0 and variance 1. Design estimates in dashed line and geostatistical estimates in full line.



→ Conclusões

Vantagens do modelo proposto:

- Considera as fontes de variabilidade mais importantes: padrões espaciais e estrutura etária.
- Permite a utilização de metodologias adaptadas à natureza das variáveis.
- Modelo paramétrico que permite a utilização de simulação estocástica para explorar características do modelo e produzir dados de entrada para ambientes de simulação com a Avaliação de Estratégias de Gestão (MSE).
- Reduz o problema da sobredispersão dos dados de campanhas, causado por muitas observações nulas e algumas observações muito elevadas:
 - O modelo espacial usa dados agregados => menos observações nulas
 - Observações nulas em idades pouco representadas são tratadas pelo modelo composicional, onde têm menos efeito nos resultados.

Desvantagens do modelo:

- Estatisticamente complexo e exigente.
- Pressuposto de independência entre Y e P.
- Revisão dos valores dos índices com a integração de um novo ano.

Obrigado !

Ernesto Jardim (IPIMAR)

[<ernesto@ipimar.pt>](mailto:ernesto@ipimar.pt)