

Styve da Conceição Alves

# ESTIMATIVA E DIAGNÓSTICO DA QUALIDADE DE AR OBTIDA POR DADOS DE OBSERVAÇÃO DA TERRA

Mestrado em Química  
Departamento de Química  
FCTUC

Setembro de 2017



UNIVERSIDADE DE COIMBRA



Styve da Conceição Alves

# **Estimativa e diagnóstico da qualidade de ar obtida por dados de observação da Terra**

**Dissertação apresentada para provas de Mestrado em Química**

Prof. Dr. Jorge Costa Pereira

Dr. Pedro Jorge Caridade

**Setembro de 2017**

**Universidade de Coimbra**

“We’ve arranged a civilization in which most crucial elements profoundly  
depend on science and technology. “

Carl Sagan

## Agradecimentos

A realização do presente trabalho tornou-se possível devido ao empenho, dedicação e amizade de algumas pessoas que contribuíram para a sua concretização.

O espaço limitado desta secção de agradecimentos, seguramente, não me permite agradecer, como devia, a todas as pessoas que, ao longo do meu Mestrado em Química me ajudaram, directa ou indirectamente, a cumprir os meus objetivos e a realizar mais esta etapa da minha formação académica. Desta forma, deixo apenas algumas palavras, poucas, mas um sentido e profundo sentimento de reconhecido agradecimento.

Assim, começo por expressar um especial agradecimento ao Professor Doutor Jorge Costa Pereira em virtude do seu excelente trabalho de orientação e supervisão desta tese, assim como, pela sua valiosa amizade e pelas inúmeras sugestões e incentivos fornecidos no decurso deste trabalho. Devo ainda o meu sincero e profundo agradecimento ao Doutor Pedro Jorge Caridade pela sua amizade e orientação desta tese, bem como, pela possibilidade de poder participar neste projeto, que muito contribuiu para o enriquecimento da minha formação académica e científica, sem dúvida, muito estimularam o meu desejo de querer, sempre, saber mais e a vontade constante de querer fazer melhor.

Devo ainda um merecido agradecimento à Primelayer e Spacelayer, a toda a sua equipa, pela colaboração no projecto, pelos dados e informações de qualidade do ar fornecidos.

Ao departamento de Química, a todos os docentes que acompanharam no meu percurso e a toda a valiosa informação transmitida a nível académico, bem como aos colegas e amigos que aí pude fazer.

Aos meus amigos, que acompanharam ao longo dos tempos, Nuno, Claudio e Miguel.

À minha família, nomeadamente aos meus pais e meu irmão Oscar, por desde sempre me terem apoiado e incentivado a aqui chegar, a eles dedico este trabalho.

A todos um grande, obrigado!

## Objectivos

Este projecto inicialmente visava caracterizar a "pegada" química de uma indústria. Contudo, atendendo a que não foi possível obter informação relativa a dados de emissões de indústrias para correlacionar com dados de satélite, teve-se de enveredar por estabelecer padrões de emissões de poluentes através dos factores meteorológicos, neste caso utilizamos, a humidade relativa, temperatura, direcção do vento, velocidade do vento e pluviosidade, recolhidos de uma sonda in-situ, com estes dados foram seleccionados os seguintes indicadores da qualidade do ar: CO, NO, NO<sub>2</sub>, NH<sub>3</sub>, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>, PANs, NMVOCs, de o sistema Copernicus. Para estabelecer estes padrões de emissão, partimos de uma estratégia de modelar por via de análise multivariada, através de dados relativos ao mês de fevereiro de 2017 dos poluentes e dos indicadores de qualidade referidos anteriormente.

## Resumo

O ar que respiramos pode conter diversos poluentes, dependendo de diversos factores que podem contribuir à sua constituição, e de modo a precaver para situações de riscos, elevados teores, estes, podem provocar graves efeitos no ambiente e na saúde pública. Nesse sentido, o presente trabalho dirige-se ao estudo da capacidade, de modelar, através de ferramentas estatísticas, os poluentes descritos anteriormente, bem como as suas relações químicas/moleculares. Obter informação quantificada no que diz respeito à influência e interdependência cruzada das diferentes vertentes de caracterização química da qualidade do ar. Esta ideia surgiu através da colaboração da Primelayer e Spacelayer, que disponibilizou os dados relativos, ao mês de fevereiro de 2017 no Meco, município de Montemor-o-velho, esse dados foram, os factores meteorológicos, a humidade relativa, temperatura, direção do vento, velocidade do vento e pluviosidade e indicadores da qualidade do ar CO, NO, NO<sub>2</sub>, NH<sub>3</sub>, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>, PANs, NMVOCs, para caracterizar os diversos poluentes químicos. Como ferramentas foi utilizado o código SNAP do Copernicus e análise multivariacional para estabelecer padrões de emissão de poluentes.

O presente estudo evidenciou uma boa descrição (modelação) dos teores de PM<sub>2.5</sub>, NO<sub>2</sub>, CO e NMVOCs.

Palavras chave: qualidade do ar, imagens de satélite, poluentes atmosféricos, modelação multivariada, escolha de modelos, descrição do processo

## Abstract

The air we breathe may contain several pollutants, depending on several factors that may contribute to its constitution, and in order to prevent risks, high levels of these, can have serious effects on the environment and public health. In this sense, the present work is directed to the study of the modeling capacity, through statistical tools, the pollutants described previously, as well as their chemical / molecular relations. To obtain quantified information regarding the influence and interdependence of the different aspects of chemical characterization of air quality. This idea arose through the collaboration of Primelayer and Spacelayer, who provided the data related to the month of February 2017 in Meco, a municipality of Montemor-o-velho. These data were, meteorological factors, relative humidity, temperature, direction of wind, wind speed and rainfall and air quality indicators CO, NO, NO<sub>2</sub>, NH<sub>3</sub>, SO<sub>2</sub>, O<sub>3</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>, PANs, NMVOCs, to characterize the various chemical pollutants. Copernicus SNAP code and multivariate analysis were used as tools to establish pollutant emission standards.

The present study evidenced a good description (modeling) of the contents of PM<sub>2.5</sub>, NO<sub>2</sub>, CO e NMVOCs.

Keywords: air quality, satellite images, atmospheric pollutants, multivariate modeling, model choice, process description



## Preâmbulo

O corpo desta tese encontra-se organizada em cinco capítulos e anexos.

No primeiro capítulo apresenta-se em linhas gerais o tópico da qualidade do ar e alguns meios disponíveis para a sua avaliação, de modo remoto, bem como as empresa (Primelayer) e a (Spacelayer) que é uma é spin-off, divisão, da Primelayer e da Universidade de Coimbra cujo principal objetivo é desenvolver uma empresa internacional dedicada à detecção do meio ambiente, com um forte laboratório de pesquisa e desenvolvimento.

Já no segundo capítulo procurou-se transmitir as bases de algumas técnicas de análise de estatística aplicadas no decorrer do trabalho.

O capítulo três trata da parte processual relacionada com a aquisição dos dados, sua sistematização, pré-processamento para posterior análise.

No capítulo seguinte apresentam-se os resultados mais significativos obtidos na análise de dez indicadores da qualidade do ar, começando por uma prévia avaliação global das variáveis e depois seguida da sua modelação multivariada, sempre que possível, apresentando as devidas observações e explicações.

No capítulo quinto reúne-se as conclusões mais significativas retiradas ao longo deste trabalho.

Por uma questão de tornar este texto de leitura mais fácil e interessante optou-se por apresentar apenas duas análises completas de resultados tendo-se remetido a informação mais pesada para anexo, no caso das restantes oito variáveis.

No sentido de uniformizar a escrita de números reais e facilitar a sua transferência entre os diversos programas utilizados optou-se por utilizar o ponto (.) como separador decimal.

# Nomenclatura

$\alpha$  - probabilidade de aceitação da hipótese nula

$\mu_i$  - estimativa central paramétrica

$\sigma_{fit}$  - desvio padrão do ajuste do modelo ( $\sigma_{fit} = \sqrt{\frac{SS}{(n-p)}}$ ) sendo  $SS$  a soma de quadrados dos desvios do modelo aos valores experimentais,  $n$  o número de pontos ajustados e  $p$  o respectivo número de parâmetros do modelo

$n$  - número de pontos a ajustar (modelar)

$p$  - número de parâmetros do modelo

$PM_{10}$  - Material particulado com diâmetro inferior a 10  $\mu\text{m}$

$PM_{2.5}$  - Material particulado com diâmetro inferior a 2.5  $\mu\text{m}$

$SS$  - soma de quadrados dos desvios entre os valores a modelar e os valores previstos pelo modelo

$X_a$  - percentil  $a$  (valor da variável que contém a percentagem de elementos)

AIC - critério de Akaike

AIC-B - critério de Akaike em modo simplificativo (Backward)

AIC-BF - critério de Akaike em modo bidirecional (primeiro em modo simplificativo, Backward, seguido de modo inverso, Forward)

BIC - critério Bayesiano (suprime um pouco o efeito da variação do número de parâmetros)

BIC-B - critério bayesiano em modo simplificativo (Backward)

BIC-BF - critério Bayesiano em modo bidirecional (primeiro em modo simplificativo, Backward, seguido de modo inverso, Forward)

CAMS - Copernicus Atmosphere Monitoring Service

COV - Compostos orgânicos voláteis

ESA - Agência Espacial Europeia

NMVOCs - compostos orgânicos voláteis não metânicos (ex: benzeno, etanol, formaldeído,...)

OMS - Organização Mundial de ou da Saúde

$p[H_0]$  - valor prova

PANs - peroxi-acetilo de nitrato ( $C_2H_3NO_5$ ), poluente secundário característico do smog

SAPALDIA - Swiss cohort study on Air Pollution And Lung and heart Diseases In Adults

SE - Erro padrão (estimativa da dispersão do parâmetro)

SOUL - Sensor Observation of Urban Life projeto da Primelayer e Spacelayer

x - variável dependente (preditores)

y - variável independente (resposta)

# Conteúdo

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introdução</b>   | <b>1</b>  |
| 1.1      | O ar, um bem essencial para a vida. . . . .                 | 1         |
| 1.2      | Saúde e poluição atmosférica, qual a sua ligação? . . . . . | 2         |
| 1.3      | Parâmetros de qualidade de ar . . . . .                     | 4         |
| 1.3.1    | Dióxido de enxofre . . . . .                                | 5         |
| 1.3.2    | Óxidos de azoto . . . . .                                   | 5         |
| 1.3.3    | Material particulado . . . . .                              | 6         |
| 1.3.4    | Ozono . . . . .   | 6         |
| 1.3.5    | Amônia . . . . .  | 7         |
| 1.3.6    | Compostos orgânicos voláteis não metânicos . . . . .        | 8         |
| 1.3.7    | Monóxido de carbono . . . . .                               | 8         |
| 1.3.8    | PANs . . . . .  | 9         |
| 1.4      | Monitorização da poluição do ar . . . . .                   | 9         |
| 1.5      | Recolha e observação de dados da Terra . . . . .            | 11        |
| <b>2</b> | <b>Fundamentação</b>  | <b>16</b> |
| 2.1      | Covariância e correlação . . . . .                          | 16        |
| 2.1.1    | Correlação . . . . .  | 17        |
| 2.2      | Regressão linear simples . . . . .                          | 18        |
| 2.2.1    | Conformidade do modelo de regressão . . . . .               | 19        |
| 2.2.1.1  | Análise residual . . . . .                                  | 19        |
| 2.3      | Modelos de regressão polinomial . . . . .                   | 20        |
| 2.3.1    | Seleção de variáveis e construção de modelos . . . . .      | 21        |
| 2.3.1.1  | Todas as regressões possíveis . . . . .                     | 22        |
| 2.3.1.2  | Seleção de modelos por etapas . . . . .                     | 22        |
| 2.3.2    | Critérios . . . . .   | 23        |
| 2.3.2.1  | Mínimos Quadrados . . . . .                                 | 24        |
| 2.3.2.2  | Coefficiente de determinação ( $R^2$ ) . . . . .            | 24        |
| 2.3.2.3  | AIC . . . . .   | 25        |
| 2.3.2.4  | BIC . . . . .   | 26        |
| <b>3</b> | <b>Experimental</b>   | <b>27</b> |
| 3.1      | Sistematização de processos . . . . .                       | 27        |
| 3.2      | Organização da informação . . . . .                         | 28        |
| 3.3      | Pré-tratamento das variáveis . . . . .                      | 28        |
| 3.4      | Modelação . . . . .   | 29        |

|          |                                  |           |
|----------|----------------------------------|-----------|
| <b>4</b> | <b>Resultados</b>                | <b>30</b> |
| 4.1      | Diagnóstico preliminar . . . . . | 30        |
| 4.1.1    | Monóxido de carbono . . . . .    | 32        |
| 4.1.2    | Dióxido de azoto . . . . .       | 33        |
| 4.1.3    | Ozono . . . . .                  | 35        |
| 4.1.4    | PM <sub>2.5</sub> . . . . .      | 36        |
| 4.1.5    | PM <sub>10</sub> . . . . .       | 38        |
| 4.1.6    | Dióxido de enxofre . . . . .     | 39        |
| 4.1.7    | Óxido de azoto . . . . .         | 41        |
| 4.1.8    | Amoníaco . . . . .               | 42        |
| 4.1.9    | NMVOCs . . . . .                 | 43        |
| 4.1.10   | PANs . . . . .                   | 44        |
| 4.2      | Modelação multivariada . . . . . | 46        |
| 4.2.1    | Monóxido de carbono . . . . .    | 46        |
| 4.2.1.1  | AIC-B . . . . .                  | 48        |
| 4.2.1.2  | AIC-BF . . . . .                 | 51        |
| 4.2.1.3  | BIC-B . . . . .                  | 55        |
| 4.2.1.4  | BIC-BF . . . . .                 | 58        |
| 4.2.1.5  | Interpretação . . . . .          | 62        |
| 4.2.2    | Dióxido de azoto . . . . .       | 64        |
| 4.2.2.1  | AIC-B . . . . .                  | 65        |
| 4.2.2.2  | AIC-BF . . . . .                 | 69        |
| 4.2.2.3  | BIC-B . . . . .                  | 73        |
| 4.2.2.4  | BIC-BF . . . . .                 | 76        |
| 4.2.2.5  | Interpretação . . . . .          | 79        |
| 4.2.3    | Ozono . . . . .                  | 80        |
| 4.2.4    | PM <sub>2.5</sub> . . . . .      | 81        |
| 4.2.5    | PM <sub>10</sub> . . . . .       | 82        |
| 4.2.6    | Dióxido de enxofre . . . . .     | 84        |
| 4.2.7    | Monóxido de azoto . . . . .      | 85        |
| 4.2.8    | Amoníaco . . . . .               | 85        |
| 4.2.9    | NMVOCs . . . . .                 | 86        |
| 4.2.10   | PANs . . . . .                   | 88        |
| <b>5</b> | <b>Conclusões</b>                | <b>90</b> |
|          | <b>Bibliografia</b>              | <b>92</b> |
| <b>A</b> | <b>Material suplementar</b>      | <b>97</b> |
| A.1      | Análise multivariada . . . . .   | 97        |
| A.1.1    | Ozono . . . . .                  | 97        |
| A.1.1.1  | AIC-B . . . . .                  | 97        |
| A.1.1.2  | AIC-BF . . . . .                 | 99        |
| A.1.1.3  | BIC-B . . . . .                  | 101       |
| A.1.1.4  | BIC-BF . . . . .                 | 102       |
| A.1.2    | PM <sub>2.5</sub> . . . . .      | 104       |
| A.1.2.1  | AIC-B . . . . .                  | 104       |
| A.1.2.2  | AIC-BF . . . . .                 | 105       |
| A.1.2.3  | BIC-B . . . . .                  | 107       |
| A.1.2.4  | BIC-BF . . . . .                 | 108       |

|       |                              |     |
|-------|------------------------------|-----|
| A.1.3 | PM <sub>10</sub> . . . . .   | 110 |
|       | A.1.3.1 AIC-B . . . . .      | 110 |
|       | A.1.3.2 AIC-BF . . . . .     | 111 |
|       | A.1.3.3 BIC-B . . . . .      | 113 |
|       | A.1.3.4 BIC-BF . . . . .     | 114 |
| A.1.4 | Dióxido de enxofre . . . . . | 116 |
|       | A.1.4.1 AIC-B . . . . .      | 116 |
|       | A.1.4.2 AIC-BF . . . . .     | 117 |
|       | A.1.4.3 BIC-B . . . . .      | 119 |
|       | A.1.4.4 BIC-BF . . . . .     | 120 |
| A.1.5 | Monóxido de azoto . . . . .  | 122 |
|       | A.1.5.1 AIC-B . . . . .      | 122 |
|       | A.1.5.2 AIC-BF . . . . .     | 123 |
|       | A.1.5.3 BIC-B . . . . .      | 125 |
|       | A.1.5.4 BIC-BF . . . . .     | 126 |
| A.1.6 | Amoníaco . . . . .           | 128 |
|       | A.1.6.1 AIC-B . . . . .      | 128 |
|       | A.1.6.2 AIC-BF . . . . .     | 129 |
|       | A.1.6.3 BIC-B . . . . .      | 131 |
|       | A.1.6.4 BIC-BF . . . . .     | 132 |
| A.1.7 | NMVOCs . . . . .             | 134 |
|       | A.1.7.1 AIC-B . . . . .      | 134 |
|       | A.1.7.2 AIC-BF . . . . .     | 135 |
|       | A.1.7.3 BIC-B . . . . .      | 137 |
|       | A.1.7.4 BIC-BF . . . . .     | 138 |
| A.1.8 | PANs . . . . .               | 140 |
|       | A.1.8.1 AIC-B . . . . .      | 140 |
|       | A.1.8.2 AIC-BF . . . . .     | 141 |
|       | A.1.8.3 BIC-B . . . . .      | 143 |
|       | A.1.8.4 BIC-BF . . . . .     | 144 |

# Capítulo 1

## Introdução

### 1.1 O ar, um bem essencial para a vida.

O ar é uma mistura de gases constituído em parte por 78% de azoto e 21% de oxigênio [1], ele é necessário para toda a vida, a falta de oxigênio leva à morte em poucos minutos. Por isso, é um pouco surpreendente ver como a humanidade negligencia a qualidade do ar que respira.

Os outros componentes são dióxido de carbono, gases raros e, dependendo da localização, uma maior ou menor quantidade de poluentes atmosféricos prejudiciais para a nossa saúde. Estes incluem material particulado ( $PM_{2.5}$ ,  $PM_{10}$ ), ozono e óxidos de azoto [1]. A poluição do ar refere-se à introdução de materiais físico-químicos ou biológicos na atmosfera que podem causar danos ou desconforto para os seres humanos ou outros organismos vivos, ou a deterioração do ambiente natural. Embora a poluição do ar tenha diminuído consideravelmente nos últimos 30 anos, o ar está longe de ser considerado limpo. Apesar de passarmos a maior parte do tempo em espaços fechados, não ao ar livre, necessitamos de uma boa qualidade de ar exterior para renovar esses espaços. Cozinhar, queimar velas ou fumar são todas atividades que influenciam a qualidade do ar interior, bem como poluentes externos.

As emissões dos principais poluentes atmosféricos na Europa diminuíram desde 1990. Durante a última década, esta redução das emissões resultou, para alguns dos poluentes, na melhoria da qualidade do ar em toda a região. No entanto, devido às complexas ligações entre as emissões e a qualidade do ar, as reduções de emissões nem sempre produzem uma correspondente queda nas concentrações atmosféricas, especialmente para PM e  $O_3$ .

Por exemplo, embora reduções de substâncias formadoras de  $O_3$  (isto é, gases precursores de  $O_3$ ) tenham sido substanciais na Europa, as concentrações de  $O_3$  na Europa permaneceram estáveis [2]. Os níveis de concentração dependem de variações nas condições climáticas, incluindo a luz solar, emissões naturais de  $O_3$ , o aumento das concentrações globais de ozono de fundo, o transporte de ozono e de substâncias precursoras de ozono principalmente  $NO_x$  (óxidos de azoto) e NMVOCs (compostos orgânicos voláteis não metânicos), mas também CO (monóxido de carbono), provenientes de áreas de origem fora da Europa. Todos estes factores contribuem para a formação de ozono, apesar das reduções europeias de emissões de poluentes, estas, podem não resultar em reduções

equivalentes das concentrações de ozono [3].

Estes números não mostram mudanças significativas ao longo dos anos [4].

Como consequência de atividades e intervenções antropogênicas sempre crescentes, nós humanos influenciemos muito os diferentes componentes do meio ambiente (por exemplo, recursos de ar, água e terra) e sofreremos os efeitos em troca.

## 1.2 Saúde e poluição atmosférica, qual a sua ligação?

Todos os dias, respiramos cerca de 15 mil litros de ar. As substâncias poluentes entram nos pulmões através do nariz, garganta e traqueia, além de oxigênio vital e colocam a saúde em risco.

Foi durante o episódio de smog [5], ou, muitas vezes chamado nevoeiro fotoquímico constituído por poluentes primários [6], de Londres em 1952 figura 1.1 que o vínculo entre poluição do ar e saúde foi estabelecido e analisado: muitas pessoas que sofrem de problemas respiratórios tiveram que ir às urgência e quase 4.000 pessoas perderam a vida como resultado dessa poluição do ar [7].



Figura 1.1: Smog em Piccadilly Circus, Londres (1952) [8]

Sabemos agora que os efeitos sobre as vias respiratórias não são as únicas consequências para a saúde. Crianças, idosos e aqueles que já sofrem de doenças respiratórias ou cardiovasculares são particularmente afetados.

Segundo o estudo SAPALDIA (Swiss cohort study on Air Pollution And Lung and heart Diseases In Adults) [9], reconhecido pela OMS e o Conselho Europeu de Investigação como um dos poucos estudos de grupos com base na população na Europa. Com estas referências, colocam numa posição relevante para estudar



questões críticas sobre a epidemiologia da poluição do ar e questões de saúde ambiental para os próximos anos. Este estudo é elaborado em 8 regiões Suíças, representativas das diversas condições ambientais, meteorológicas e sociodemográficas deste país.

O estudo SAPALDIA examinou os vínculos entre qualidade do ar e saúde. Ele demonstrou que as pessoas expostas a uma maior quantidade de poluentes apresentam maior risco de contrair diferentes doenças ou apresentar certos sintomas. Embora as concentrações de poluentes sejam relativamente baixas, como é o caso na Suíça, podem ser esperados efeitos na saúde. O estudo também descobriu que a melhoria da qualidade do ar retarda a degradação da função pulmonar relacionada à idade e diminui os sintomas respiratórios [10].

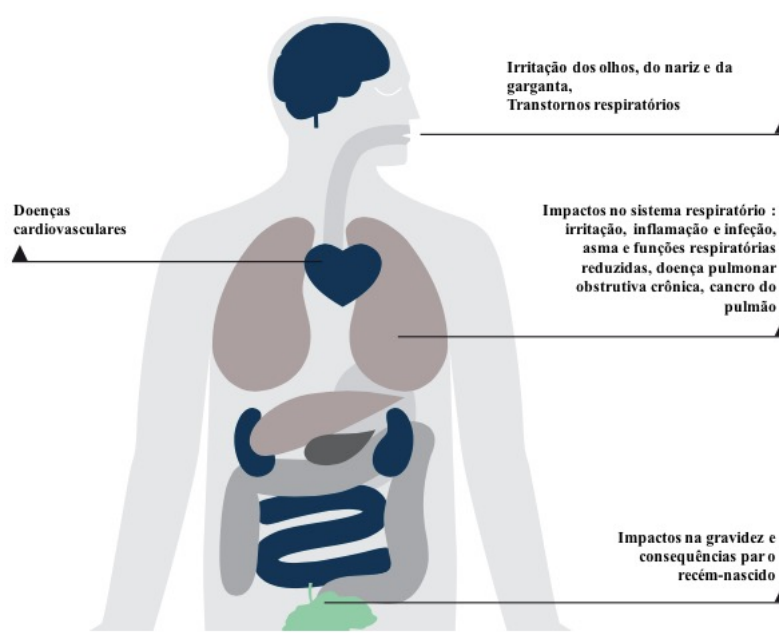


Figura 1.2: Efeitos de a poluição atmosférica sobre a saúde.

A inalação de poluentes do ar pode irritar as vias respiratórias, causar sintomas respiratórios e levar a doenças agudas ou crônicas. As partículas de poeira fina com um diâmetro de cerca de 10  $\mu\text{m}$  atingem os brônquios. As partículas mais finas e os gases como o ozono ou os óxidos de azoto atingem as vias respiratórias mais profundas, os alvéolos. Os poluentes desencadeiam reações inflamatórias, que podem ser manifestadas por tosse e expectoração [11].

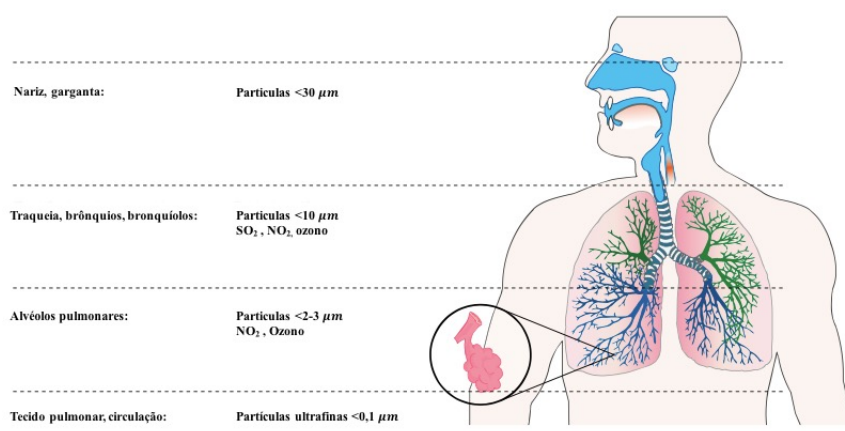


Figura 1.3: Profundidade de penetração de poluentes nas vias respiratórias. [12]

A gravidade das consequências da poluição do ar para as vias respiratórias depende da concentração de poluentes no ar, da duração da exposição e da taxa respiratória, entramos nos principais tópicos em que a detecção remota pode ter um papel relevante e eficaz, ao alertar para situações críticas ajudando a minimizar os efeitos nefastos da poluição no ambiente e principalmente na saúde pública.

### 1.3 Parâmetros de qualidade de ar

A qualidade do ar depende das emissões dos poluentes mas está também sujeita à variabilidade dos fenômenos atmosféricos, os quais desempenham um papel preponderante nos processos de transporte, transformação e dispersão dos poluentes na atmosfera. Estes processos são influenciados pela topografia local e por factores meteorológicos como o vento, a pressão atmosférica, a temperatura, a precipitação e a radiação solar [13].

O vento é um fator meteorológico com efeitos diretos e determinantes nas condições de dispersão dos poluentes. A velocidade do vento determina a produção de turbulência mecânica, que é responsável pela dispersão local. A ausência de vento favorece a concentração de poluentes, e situações de vento moderado favorecem a sua dispersão, no entanto, o vento forte pode provocar um efeito de penacho e poluição localizada na direção dos ventos dominantes.

A temperatura intervém na química dos poluentes e desempenha também um papel importante na sua dispersão vertical na atmosfera. No verão temperaturas elevadas favorecem a formação do ozono e no inverno as diferenças de temperatura entre o dia e a noite podem provocar inversões térmicas e picos de poluição. A estabilidade atmosférica determina os processos convectivos locais sendo caracterizada pelo gradiente vertical de temperatura que limita a mistura vertical de poluentes se existir uma inversão térmica. A temperatura do ar tende a diminuir em altura, no entanto, em determinadas condições pode ocorrer uma inversão térmica, ou seja, pode verificar-se um aumento de temperatura, criando uma camada de ar quente que impede o ar poluído junto ao solo de subir e se dispersar.

A precipitação está geralmente associada a uma atmosfera instável, favorecendo uma boa dispersão dos poluentes atmosféricos. As gotas de chuva solubilizam os poluentes gasosos e as partículas, provocando a sua deposição sobre o solo e outras superfícies, diminuindo assim as concentrações no ar ambiente

O objectivo é estabelecer padrões de emissões de poluentes através dos factores meteorológicos, neste caso utilizamos, a humidade relativa, temperatura, direção do vento, velocidade do vento e pluviosidade, para tal, foram seleccionados os seguintes indicadores da qualidade do ar, bem como as suas características e propriedades [4].

### 1.3.1 Dióxido de enxofre

O dióxido de enxofre ( $\text{SO}_2$ ) é formado pela oxidação do enxofre (S), principalmente através da combustão de combustíveis contendo (S). O setor de produção de electricidade é a fonte mais importante de  $\text{SO}_2$ . O  $\text{SO}_2$  também pode contribuir para a formação de partículas secundárias de sulfato na atmosfera [14]. É um gás irritante para as mucosas oculares e vias respiratórias podendo ter efeitos agudos ou crónicos na saúde humana, essencialmente no aparelho respiratório. Agrava problemas cardiovasculares devido ao seu impacto na função respiratória. O dióxido de enxofre resultante da queima de combustíveis pode transformar-se em trióxido de enxofre ( $\text{SO}_3$ ) que, na presença de humidade atmosférica, originará ácido sulfúrico ( $\text{H}_2\text{SO}_4$ ) e seus sais [14]. Os seus efeitos nas plantas consistem na alteração do metabolismo e diminuição da taxa de crescimento, principalmente quando sujeito a temperaturas reduzidas. Outros efeitos são o necrosamento de tecidos e aumento da sensibilidade a temperaturas baixas e aos parasitas. A deposição de dióxido de enxofre e de aerossóis sulfurados sobre edificações e materiais de construção acelera a sua corrosão e envelhecimento .

### 1.3.2 Óxidos de azoto

Os óxidos de azoto ( $\text{NO}_x$ ) são emitidos durante a combustão de combustível, principais emissores, instalações industriais e dos transportes rodoviários. Os compostos de azoto mais importantes em termos de poluição atmosférica são monóxido de azoto (NO) e dióxido de azoto ( $\text{NO}_2$ ). O dióxido de azoto é um gás facilmente detectável pelo odor característico, de cor acastanhada, corrosivo e extremamente oxidante. O monóxido de azoto é um gás incolor e inodoro, não tóxico para as concentrações habitualmente presentes na atmosfera. Tanto o monóxido como o dióxido de azoto, que em conjunto são designados  $\text{NO}_x$ , têm origem natural, resultante do metabolismo microbiano dos solos e de descargas eléctricas na atmosfera (durante as trovoadas) [15], e também antropogénica, resultante da queima de combustíveis a altas temperaturas, quer em instalações

fixas, quer em veículos automóveis (principal fonte deste poluente nas áreas urbanas). Quando as condições são favoráveis o monóxido de azoto emitido para a atmosfera oxida-se fotoquimicamente originando o respetivo dióxido. Quando as condições meteorológicas são favoráveis (radiação solar e temperatura elevadas) os óxidos de azoto podem reagir com os compostos voláteis originando, entre outros, o ozono ( $O_3$ ) [16], composto oxidante fotoquímico muito forte. O  $NO_x$  contribui para a formação de ozono, partículas e pode reagir com a amônia para formar partículas de ácido nítrico

### 1.3.3 Material particulado

O material particulado (PM) é um conjunto complexo de substâncias, minerais ou orgânicas, que se encontram em suspensão na atmosfera, sob a forma líquida ou sólida [17]. A sua dimensão pode variar entre algumas dezenas de nanómetros e uma centena de micrómetros ( $\mu m$ ). Partículas finas com um diâmetro entre 2,5 e 10 micrómetros ( $\mu m$ ) ( $PM_{10}$ ), partículas extrafinas com um diâmetro de 2,5  $\mu m$  ou menos ( $PM_{2.5}$ ). PM é emitido diretamente como partículas primárias ou formadas na atmosfera através de emissões de  $SO_2$ ,  $NO_x$ ,  $NH_3$  e NMVOCs. O PM são emitidos para a atmosfera a partir de uma gama variada de fontes antropogénicas sendo as mais importantes a queima de combustíveis fósseis, o tráfego rodoviário e determinados processos industriais. Estas substâncias podem também ser emitidas por fontes naturais tais como os vulcões, fogos florestais ou serem resultantes da ação do vento sobre o solo e superfícies aquáticas. Na bacia Mediterrânica e nos arquipélagos do Atlântico, os eventos naturais de intrusão de massas de ar com partículas em suspensão com origem nos desertos do Norte de África são uma fonte natural importante de partículas. Em Portugal, este fenómeno ocorre também com alguma frequência. A exposição crónica a partículas contribui para o risco de desenvolvimento de doenças respiratórias e cardiovasculares [18], assim como para o cancro de pulmão [19]. As partículas em suspensão são também um veículo de transporte eficaz para outros poluentes atmosféricos que se fixam à sua superfície, especialmente hidrocarbonetos e metais pesados. Estas substâncias são muitas vezes transportadas até aos pulmões onde podem depois ser absorvidas para o sangue e tecidos.

### 1.3.4 Ozono

O ozono ( $O_3$ ) do nível do solo (troposférico) não é emitido diretamente para a atmosfera. Em vez disso, ele é formado na atmosfera a partir de uma cadeia de reações químicas após as emissões de certos gases precursores [20]:  $NO_x$ , monóxido de carbono (CO), NMVOCs e metano ( $CH_4$ ). O ozono é um gás essencial na estratosfera, na medida em que possui a capacidade de limitar a radiação ultravioleta que atinge a superfície terrestre [21]. No entanto, na troposfera é considerado um dos poluentes mais nocivos para a saúde. Este poluente é formado por reação lenta entre os óxidos de azoto e voláteis, por ação da radiação solar. Os compostos orgânicos voláteis são emitidos nas combustões

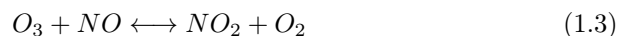
incompletas e por volatilização de combustíveis. A formação deste poluente é lenta (período de formação de várias horas) e a sua destruição demora apenas alguns minutos. O processo de formação inicia-se com a fotólise do dióxido de azoto:



O oxigénio atómico é altamente reactivo pelo que reage com o oxigénio diatómico formando ozono segundo a reacção:



O ozono assim formado é rapidamente destruído por reacção com o monóxido de azoto:



O monóxido de azoto é regenerado por esta reacção e, na ausência de COVs é obtido um estado de equilíbrio dinâmico entre as reacções de formação e destruição de ozono. Nas zonas não diretamente influenciadas pelas emissões de tráfego automóvel o equilíbrio referido é atingido durante o dia, quando a radiação solar provoca a fotólise do  $\text{NO}_2$  com formação de Ozono. Durante a noite, na ausência da radiação solar necessária para iniciar a primeira fase do ciclo, o ozono é destruído sem ser repostado, formando-se uma reserva de  $\text{NO}_2$  que poderá originar ozono no dia seguinte. Este ciclo básico não permite por si só explicar a formação de ozono na atmosfera urbana. Neste caso contribuem activamente outros oxidantes fotoquímicos formados a partir dos COVs, que possuem a capacidade de oxidar o NO em  $\text{NO}_2$  sem consumo de ozono. O NO ao reagir com estes oxidantes fica menos disponível para destruir o ozono. Os COVs aceleram a oxidação do NO a  $\text{NO}_2$ , que, por sua vez reage na presença de radiação para formar ozono. Esta reacção em cadeia permite a formação de um excesso de ozono mesmo quando as concentrações dos seus precursores são relativamente reduzidas. O ozono penetra profundamente nas vias respiratórias, afetando essencialmente os brônquios e os alvéolos pulmonares, fazendo a sua ação sentir-se mesmo em concentrações baixas e em exposições de curta duração, principalmente em crianças e asmáticos, manifestando-se, inicialmente, por irritação dos olhos, nariz e garganta, seguindo-se tosse e dor de cabeça. Os efeitos nocivos deste poluente são acentuados com a actividade física intensa, por aumento da taxa respiratória. O ozono provoca também danos nas espécies vegetais nomeadamente, manchas nas folhas, redução do crescimento e decréscimo de produtividade [22]. Os danos provocados pelo ozono em materiais como a borracha e pigmentos podem também ser elevados.

### 1.3.5 Amônia

A amônia ( $\text{NH}_3$ ), a grande maioria das emissões de  $\text{NH}_3$  provém do setor agrícola, em relação a atividades como armazenamento de dejetos, espalhamento de lamas e uso de fertilizantes nitrogenados sintéticos [17]. Contribui também

para a formação de partículas.

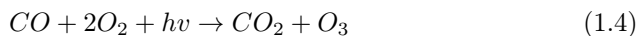
### 1.3.6 Compostos orgânicos voláteis não metânicos

Os compostos orgânicos voláteis não metânicos (NMVOCs), são uma grande variedade de compostos quimicamente diferentes, tais como benzeno, etanol, formaldeído, ciclo-hexano, 1,1,1-tricloroetano ou acetona. Produzem oxidantes fotoquímicos por reação com NO<sub>x</sub> na presença de luz solar. Os NMVOCs antropogênicos são emitidos a partir de fontes, incluindo aplicação de tinta, transporte rodoviário, limpeza a seco e outros usos de solventes. Os NMVOCs biogênicos são emitidos pela vegetação, com as quantidades emitidas dependentes das espécies e da temperatura.

### 1.3.7 Monóxido de carbono

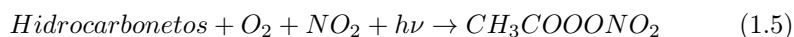
O monóxido de carbono (CO) é um gás incolor e inodoro, que ocorre naturalmente de erupções vulcânicas, fogos florestais e da decomposição da clorofila. Fontes importantes de CO incluem, o transporte rodoviário, as empresas, as famílias e a indústria, emitido devido à combustão incompleta. Por esse motivo é considerado um bom indicador da poluição resultante do tráfego rodoviário. A sua concentração diminui rapidamente com o aumento da distância às fontes de emissão. No que respeita à saúde humana, salienta-se que a sua tendência para se fixar na hemoglobina é cerca de 210 vezes superior à do oxigénio, pelo que esta fica bloqueada na forma de carboxihemoglobina. Em caso de exposição prolongada verificam-se dificuldades respiratórias que poderão conduzir à morte. O CO reage com outros poluentes que produzem ozono no nível do solo (troposférico) [23].

O monóxido de carbono, em conjunto com aldeídos, participam em reações químicas que formam poluição fotoquímica. Reage com o radical hidroxilo (OH•) para produzir um intermediário radical (HOCO•), que transfere rapidamente o hidrogénio radical para O<sub>2</sub> para formar o radical peroxi/hidroperoxilo (HO<sub>2</sub>•) e o dióxido de carbono (CO<sub>2</sub>). [24] O radical peroxi reage posteriormente com óxido de azoto (NO) para formar dióxido de azoto (NO<sub>2</sub>) e radical hidroxilo. O NO<sub>2</sub> dá um átomo de oxigénio por via fotólise, formando assim O<sub>3</sub> após a reação com O<sub>2</sub>. Uma vez que o radical hidroxilo é formado durante a formação de NO<sub>2</sub>, o equilíbrio da sequência de reações químicas que começam com monóxido de carbono e que conduz à formação de ozono é:



### 1.3.8 PANs

Os PANs, ou, peroxi-acetilo de nitratos ( $C_2H_3NO_5$ ), são poderosos irritantes respiratórios e oculares presentes em poluição fotoquímica. São nitratos produzidos no equilíbrio químico entre radicais peroxi orgânicos pela oxidação em fase gasosa de uma variedade de compostos orgânicos voláteis, ou por aldeídos e outros COVs oxigenados que se oxidam na presença de  $NO_2$ .



Eles são bons marcadores para a fonte de COVs como biogênicos ou antropogênicos, o que é útil no estudo dos efeitos globais e locais de poluentes [25].

Os PANs são tóxicos e irritantes, já que eles são mais solúveis na água do que o ozônio. Eles são agentes lacrimogênicos, causando irritação nos olhos em concentrações de apenas algumas partes por bilhão. Em concentrações mais elevadas causam danos extensivos à vegetação. Ambos os PANs e seus derivados clorados são ditos mutagênicos, pois podem ser um fator que causa câncer da pele.

Os PANs são poluentes secundários, o que significa que eles não são emitidos diretamente de fábricas de produção de eletricidade ou motores de combustão interna, mas são formados por outros poluentes por meio de reações químicas na atmosfera. As reações de radicais livres catalisadas pela luz ultravioleta do sol oxidam hidrocarbonetos não queimados para aldeídos, cetonas e compostos de dicarbonilo, cujas reações secundárias criam radicais peroxiacil, que se combinam com dióxido de nitrogênio para formar nitratos de peroxiacil.

O radical peroxiacilico mais comum é o peroxi-acetilo, que pode ser formado a partir da oxidação de radicais livres de acetaldeído, várias cetonas ou a fotólise de compostos de dicarbonilo tais como metilglioal ( $C_3H_4O_2$ ) ou diacetilo ( $C_4H_6O_2$ ).

Uma vez que eles se dissociam bastante devagar na atmosfera em radicais e  $NO_2$ , os PANs são capazes de transportar esses compostos instáveis longe da origem urbana e industrial. Isso é importante para a produção de ozônio troposférico à medida que os PANs transportam  $NO_x$  para regiões onde ele pode produzir mais eficientemente o ozônio.

## 1.4 Monitorização da poluição do ar

A composição do ar ambiente e interior tem um impacto considerável na nossa saúde e qualidade de vida. A poluição do ar e as emissões de gases de efeito estufa podem ter um impacto considerável no meio ambiente, incluindo questões ambientais globais mais amplas, como o desaparecimento de ozônio estratosférico, a quantidade de moléculas de ozônio contidas numa coluna vertical de base unitária que se estende desde a superfície até o "topo" da Atmosfera [26], e as mudanças climáticas. Os poluentes do ar podem ser amplamente classificados como primários [6], quando são diretamente emitidos pelas fontes emissoras, estes poluentes podem sofrer transformações e reações químicas na atmosfera e dar origem a poluentes secundários [27], obtemos assim duas classificações, primários e secundários. Normalmente, os poluentes atmosféricos primários são

substâncias que são diretamente emitidas por um processo natural ou antropogênico, como cinzas de uma erupção vulcânica, gás de monóxido de carbono de escape de veículos motorizados ou dióxido de enxofre libertado de fábricas. No entanto, os poluentes primários não produzem, por si só, todos os efeitos adversos da poluição do ar. Podem ocorrer reações químicas entre os principais contaminantes e constituintes da atmosfera, conforme ilustrado na Figura 1.4.

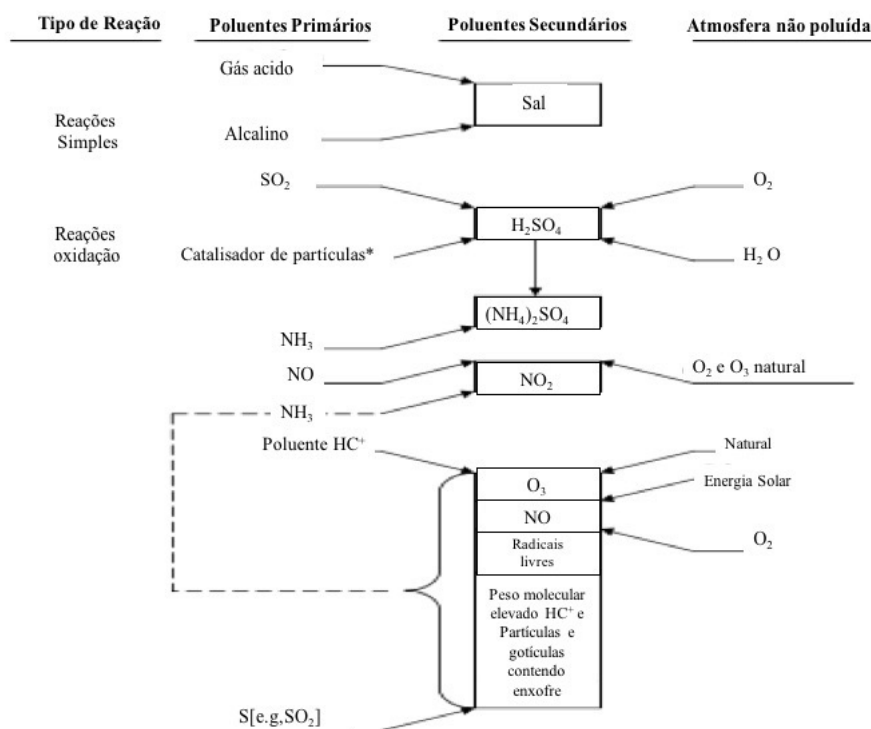


Figura 1.4: Ilustração de Poluentes primários e secundários. \* A reação pode ocorrer sem catalisador (HC +, hidrocarbonetos) [28]

Posteriormente, os poluentes gasosos que reagem uns com os outros e com partículas no ar produzem um conjunto complexo de novos compostos químicos. Os poluentes atmosféricos não emitidos diretamente como tal, mas formados no ar, são chamados de poluentes atmosféricos secundários, que são responsáveis por vários efeitos negativos da poluição do ar, como poluição atmosférica, névoa, irritação ocular e danos à vegetação e ao material. Portanto, reduzir os efeitos adversos atribuídos às exposições dos poluentes é um importante objetivo de bem-estar público visando a obtenção de benefícios sociais significativos, chegamos aos principais objetivos de monitorizar o ar [29]:

- proporcionar uma base científica sólida para o desenvolvimento de políticas e políticas de controle econômicas para reduzir a poluição do ar
- para avaliar quão longe os padrões de qualidade do ar, valores-limite e objetivos estão sendo alcançados



- avaliar potenciais impactos da poluição do ar sobre saúde e bem-estar da população
- para determinar o impacto da poluição do ar nos ecossistemas e no nosso ambiente natural
- proporcionar ao público informações confiáveis e atualizadas sobre a poluição do ar
- para cumprir os requisitos legais de relatório da qualidade do ar

No entanto, existem limitações como qualquer programa de monitoramento e uso de dados analíticos de uma técnica, portanto, deve ser usado em conjunto com os resultados de outras técnicas de avaliação, incluindo modelos computacionais, inventários de emissões, interpolação e mapeamento.

## 1.5 Recolha e observação de dados da Terra

Na abordagem do presente trabalho, as principais informações recolhidas de observação para posterior análise, foram cedidas através da colaboração com a Spacelayer e Primelayer, empresas pioneiras do sector, sediadas em Coimbra, sem o qual não seria possível a execução do projecto, pretendo realçar mais uma vez a sua colaboração e seu apoio para elaboração do trabalho.

Os dados foram recolhidos de duas fontes, satélite (CAMS) e in-situ (Montemor-o-velho), vamos começar por analisar a primeira.

O sistema Copernicus da União Europeia, ou, Copernicus Serviço de Monitorização de Atmosfera(CAMS), assenta numa parceria estabelecida entre a União Europeia (UE), a Agência Espacial Europeia (ESA) e os vários Estados-Membros (EM) é o maior programa a nível de monitoramento ambiental no mundo, anteriormente conhecido como GMES (Monitorização Global do Ambiente e da Segurança) [30].

Utiliza um sistema abrangente de monitorização e previsão global que calcula diariamente o estado da atmosfera, combinando informações de modelos, observações e fornece uma previsão diária de 5 dias, este sistema adquire dados de múltiplas fontes figura 1.5: satélites de observação da Terra e sensores in situ, como estações terrestres, sensores aéreos e marítimos.

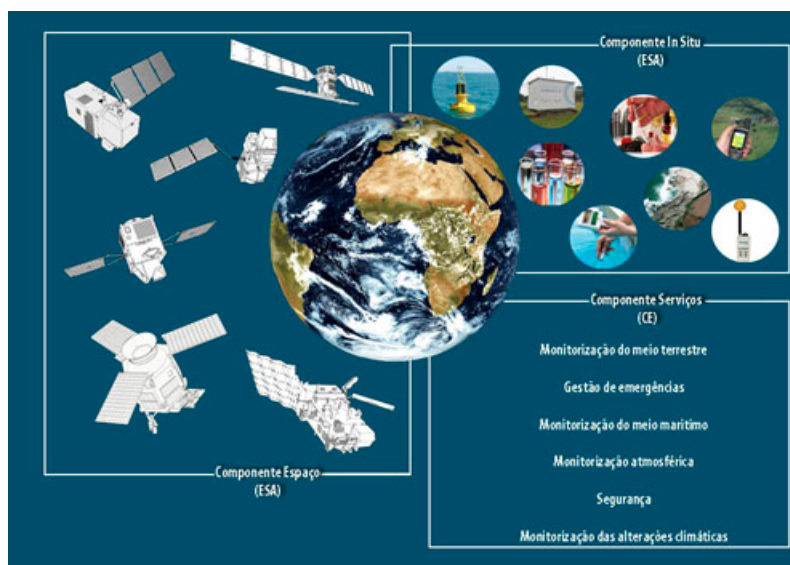


Figura 1.5: Esquematização das três componentes mais relevantes do sistema Copernicus em termos de monitorização ambiental.

O Copernicus processa esses dados e fornece aos usuários informações confiáveis e atualizadas através de um conjunto de serviços relacionados a questões ambientais e de segurança. O sistema global de modelação também é usado para fornecer as condições de limite para um conjunto de modelos de qualidade de ar regionais mais detalhados que são usados para ampliar o domínio europeu e produzir previsões de qualidade do ar de 4 dias.

Em setembro de 2017, será dado mais um grande passo através de um pioneiro e sofisticado satélite, destinado ao estudo de poluentes presentes na atmosfera é o Sentinel-5P figura 1.6,

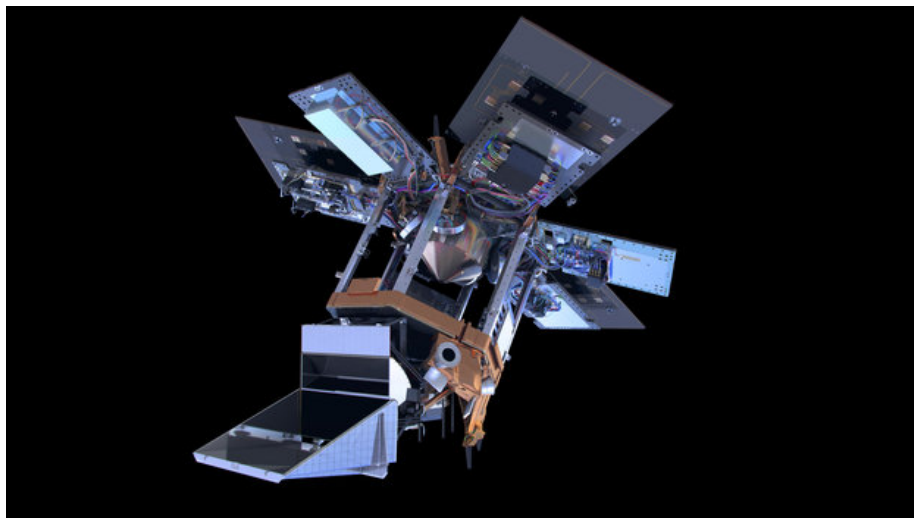


Figura 1.6: Satélite Sentinel-5P (The Copernicus Atmosphere Monitoring Service)

O Sentinel-5P possui a bordo, o Tropomi (espectrómetro) de uma única via, gama espectral ver figura 1.7, com uma faixa de alcance de 2600 km, o que permite que todo o planeta seja mapeado a cada 24 horas [31]. A plataforma do satélite é constituída em uma estrutura hexagonal e três asas solares que são activadas quando está em órbita. No seu interior abriga o sensor Tropomi e instrumentos como o conjunto óptico, o sistema de arrefecimento radiante e unidade de controle de instrumentos. O Tropomi está montado no "andar superior" da plataforma e posiciona-se num ângulo de  $30^\circ$  para garantir que o instrumento aponte para a Terra [31]. O restante equipamento, antenas e sistema de arrefecimento radiante também estão montados na parte superior. O sistema de propulsão monopropelente, gera a energia de um só composto através de produtos gasosos por intermédio de reações exotérmicas, utiliza como combustível hidrazina ( $N_2H_4$ ), é um líquido muito utilizado neste tipo de sistemas [32], com propriedades similares a amônia mas com diferente disposição espacial de seus dois átomos de hidrogénio faz com que a substância seja muito mais reactiva, tem uma capacidade de 82 kg com quatro propulsores de 1 N em dois pares redundantes. O satélite voa em uma órbita sincronizada com o Sol a 824 km acima da Terra[31].

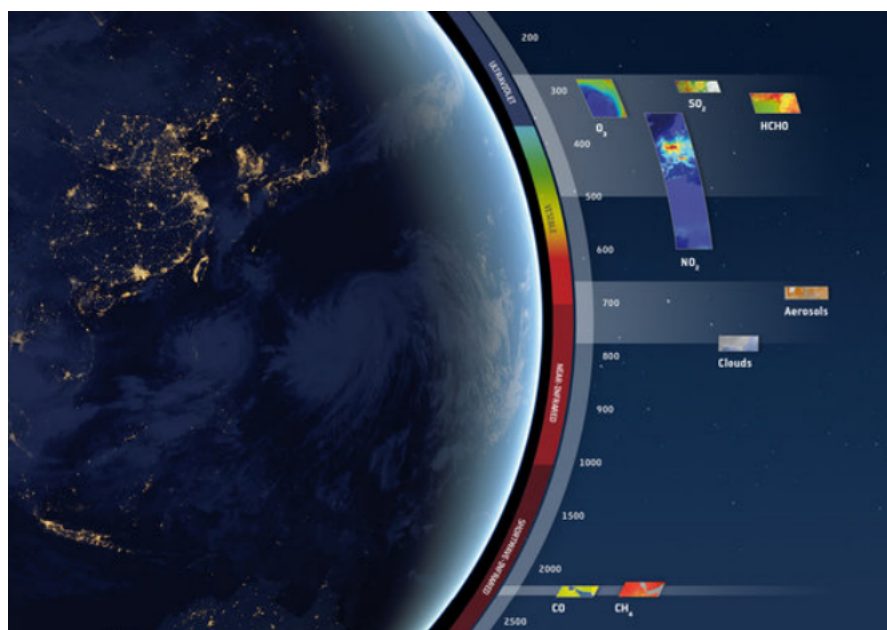


Figura 1.7: Escala espectral de Tropomi (Sentinel-5P)

O espectrômetro de imagem multiespectral incorporado (Tropomi) é o mais avançado até à data, foi desenvolvido conjuntamente pela ESA e pelo Netherlands Space Office[33]. Observa a luz solar que é espalhada de volta ao espaço pela superfície e atmosfera da Terra, detectando as impressões digitais únicas de gases em diferentes partes do espectro figura 1.7[33]. O espectro eletromagnético descreve toda a distribuição de radiação eletromagnética de acordo com a frequência ou comprimento de onda. Todas as ondas eletromagnéticas viajam com a mesma velocidade no vácuo à velocidade da luz, que é 299.792.458 metros por segundo. No entanto, a distribuição abrange uma ampla gama de frequências e comprimentos de onda, que podem ser divididas em intervalos, ou faixas do espectro eletromagnético [34]. As várias faixas possuem nomes diferentes com base em diferenças de comportamento na emissão, transmissão e absorção das ondas correspondentes e também com base em suas diferentes aplicações práticas. O que diferencia o Tropomi [35] é que ele mede nas faixas espectrais ver figura 1.7, ultravioleta e visível (270-500 nm), infravermelho próximo (675-775 nm) e infravermelho de ondas curtas (2305-2385 nm) [33].[36] Isso significa que uma ampla gama de poluentes, como dióxido de azoto, ozônio, formaldeído, dióxido de enxofre, metano e monóxido de carbono podem ser detectados com mais precisão do que nunca. Com uma resolução até  $7 \text{ km} \times 3,5 \text{ km}$ , tem potencial para detectar a poluição do ar em cidades individuais[33]. Um contributo essencial para monitorizar a qualidade do ar e fornecer informações críticas aos serviços e entidades governamentais para melhorar a vida dos cidadãos europeus. Com a cobertura global, poderá ser possível monitorizar poluição e melhorar a compreensão dos processos atmosféricos químicos e físicos.

Passando para a segunda fonte de recolha de informação, onde foram ex-

traídos os dados relativos, à humidade relativa, temperatura, direção do vento, velocidade do vento, pluviosidade da sonda meteorológica situada no Meco (fig. 1.8) no município de Montemor-o-velho. A sonda está incorporada no projeto no qual vou descrever.

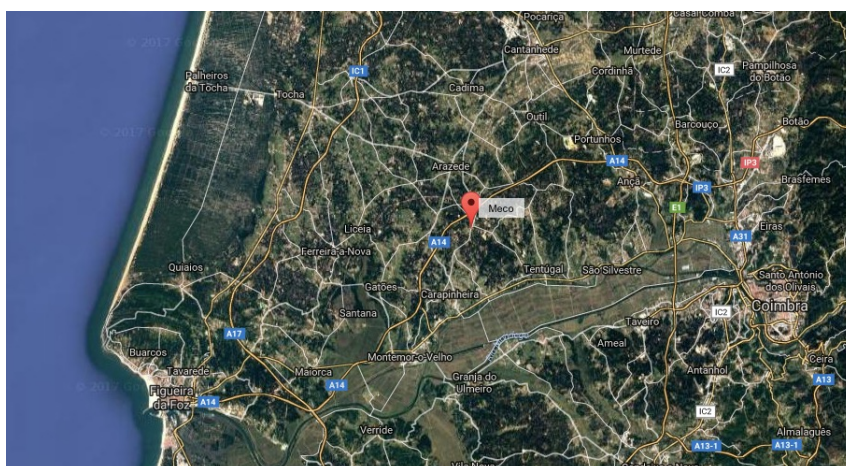


Figura 1.8: Localização da fonte in-situ de recolha de dados (sonda do projecto SOUL).

O projeto Sensor Observation of Urban Life (SOUL)[37] desenvolvido na Spacelayer e Primelayer, é uma ferramenta inteligente que ajuda a minimizar o risco de exposição à poluição do ar, a principal causa ambiental para óbitos prematuros com maior impacto na produtividade e no bem-estar. Com base em uma rede de sensores instalada em veículos em movimento, combinada com imagens de satélites e também informações personalizadas, adaptadas ao perfil médico, previsões e percepções são fornecidas aos indivíduos para permitir decisões pró-ativas que favoreçam sua aptidão e saúde.

Serão sem duvida ferramentas muito importantes no futuro, face às previsões actuais, o referente trabalho pretende abraçar estas duas tecnologias.

## Capítulo 2

# Fundamentação

Alguns dos mais importantes avanços da ciência deram-se com a modelação e a inferência estatística. Com o resultado de uma amostra, e tendo por base a formulação da relação entre uma ou mais variáveis resposta e uma ou mais variáveis predictoras, tornou-se possível a extrapolação do resultado para a população e a partir daí a descrição de uma parte do mundo real. A literatura da química analítica apresenta uma grande quantidade de publicações que se dedicam ao estudo de comparação de métodos por meio de análise de regressão. No presente capítulo, alguns dos métodos citados foram escolhidos, e pretende-se mostrar como obter de forma a modelar adequadamente as variáveis de resposta de um conjunto de regressores/predictores.

### 2.1 Covariância e correlação

A variável refere uma característica qualitativa que pode ser medida e expressa em números sob uma escala de medição contínua, quando duas ou mais variáveis aleatórias, são definidas em um espaço de probabilidade, é útil descrever como elas variam juntas; isto é, é útil medir a relação entre as variáveis. Uma medida comum da relação entre duas variáveis aleatórias é a covariância. Para definir a covariância, precisamos descrever o valor esperado de uma função de duas variáveis aleatórias  $h(x, y)$ . A definição simplesmente se estende ao usado para uma função de duas variáveis aleatórias [38].

$$E[h(x, y)] = \left\{ \sum_R \sum h(x, y) f_{xy}(x, y) \quad x, y \text{ discreto} \right. \quad (2.1)$$

Ou seja,  $E[h(x, y)]$  pode ser pensado como a média ponderada de  $h(x, y)$  para cada ponto na faixa de  $(x, y)$ . O valor de  $E[h(x, y)]$  representa o valor médio de  $h(x, y)$  que se espera em uma longa sequência de ensaios repetidos ( $\sum_R$ ) da experiência aleatória.

A Covariância entre as variáveis aleatórias  $x$  e  $y$ , denotada como  $cov(x, y)$  ou  $\sigma_{xy}$ , é [38]

$$\sigma_{xy} = E[(x - \mu_x)(y - \mu_y)] = E(xy) - \mu_x \mu_y \quad (2.2)$$

Se os pontos na distribuição de probabilidade conjunta de  $x$  e  $y$  que recebem probabilidade positiva tendem a cair ao longo de uma linha de inclinação positiva

(ou negativa),  $\sigma_{xy}$  é positivo (ou negativo). Se os pontos tendem a cair ao longo de uma linha de inclinação positiva,  $x$  tende a ser maior que  $\mu_x$  quando  $y$  é maior do que  $\mu_y$ . Portanto, o produto dos dois termos  $x - \mu_x$  e  $y - \mu_y$  tende a ser positivo. No entanto, se os pontos tendem a cair ao longo de uma linha de inclinação negativa,  $x - \mu_x$  tende a ser positivo quando  $y - \mu_y$  é negativo, ou vice-versa. O produto de  $x - \mu_x$  e  $y - \mu_y$  tende a ser negativo. Nesse sentido, a covariância entre  $x$  e  $y$  descreve a variação entre as duas variáveis aleatórias. A covariância é uma medida de relação linear entre as variáveis aleatórias. Se a relação entre as variáveis aleatórias não for linear, a covariância pode não ser sensível ao relacionamento.

### 2.1.1 Correlação

Há outra medida da relação entre duas variáveis aleatórias que é muitas vezes mais fácil de interpretar do que a covariância.

A correlação entre variáveis aleatórias  $x$  e  $y$ , denotada como  $\rho_{xy}$ , é

$$\rho_{xy} = \frac{\text{cov}(x, y)}{\sqrt{V(x)V(y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (2.3)$$

Para  $x > 0$  e  $y > 0$ , se a covariância entre  $x$  e  $y$  for positiva, negativa ou zero, a correlação entre  $x$  e  $y$  é positiva, negativa ou zero, respectivamente. O resultado a seguir pode ser mostrado.

A correlação apenas dimensiona a covariância pelo desvio padrão de cada variável. Consequentemente, a correlação é uma quantidade sem dimensão que pode ser usada para comparar as relações lineares entre pares de variáveis em diferentes unidades.

Se os pontos na distribuição de probabilidade conjunta de  $x$  e  $y$  que recebem probabilidade positiva tendem a cair ao longo de uma linha de declive positiva (ou negativa),  $\rho_{xy}$  é perto de 1 (ou -1). Se  $\rho_{xy}$  é igual a 1 ou -1, pode-se mostrar que os pontos na distribuição de probabilidade conjunta que recebem probabilidade positiva caem exatamente ao longo de uma linha reta. Duas variáveis aleatórias com correlação zero não são ditas correlacionadas. Semelhante à covariância, a correlação é uma medida da relação linear entre variáveis aleatórias.

Se  $x$  e  $y$  são variáveis aleatórias independentes,

$$\sigma_{xy} = \rho_{xy} = 0 \quad (2.4)$$

No entanto, se a correlação entre duas variáveis aleatórias é zero, não podemos imediatamente concluir que as variáveis aleatórias são independentes.

A distribuição condicional de  $y$  para um dado valor de  $x$  é [38]

$$f_{y|x}(y) = \frac{1}{\sqrt{2\pi\sigma_{y|x}}} \exp \left[ -\frac{1}{2} \left( \frac{y - \beta_0 - \beta_1 x}{\sigma_{y|x}} \right)^2 \right] \quad (2.5)$$

Onde

$$\beta_0 = \mu_y - \mu_x \rho \frac{\sigma_y}{\sigma_x} \quad (2.6)$$

$$\beta_1 = \frac{\sigma_y}{\sigma_x} \rho \quad (2.7)$$

E a variância da distribuição condicional de  $y$  dado  $x$  é

$$\sigma_{y|x}^2 = \sigma_y^2(1 - \rho^2) \quad (2.8)$$

Ou seja, a distribuição condicional de  $y$  dado  $x$  é normal com média

$$E(y | x) = \beta_0 + \beta_1 x \quad (2.9)$$

e a variância  $\sigma_{y|x}^2$ . Assim, a média da distribuição condicional de  $y$  dada  $x$  é um modelo de regressão linear simples. Além disso, existe uma relação entre o coeficiente de correlação  $\rho$  e a inclinação  $\beta_1$ . Da Equação 2.7 vemos que se  $\rho = 0$ , então  $\beta_1 = 0$ , o que implica que não há regressão de  $y$  em  $x$ . Ou seja, o conhecimento de  $x$  não nos ajuda a prever  $y$ .

## 2.2 Regressão linear simples

Sir Francis Galton utilizou pela primeira vez o termo análise de regressão em um estudo das alturas dos pais ( $x$ ) e dos filhos ( $y$ ) [38]. Galton ajustava uma linha de mínimos quadrados e usava-a para prever a altura do filho da altura dos pais. Ele descobriu que, se a altura do pai fosse acima da média, a altura do filho também seria acima da média, mas não tanto quanto o tamanho do pai. Um efeito semelhante foi observado para alturas curtas. Ou seja, a altura do filho "regrediu" em relação à média. Consequentemente, Galton referiu a linha de mínimos quadrados como uma linha de regressão.

O caso da regressão linear simples considera um único regressor ou preditor  $x$  e uma variável dependente ou de resposta  $y$ . Suponha que a relação verdadeira entre  $y$  e  $x$  seja uma linha reta e que a observação  $y$  em cada nível de  $x$  seja uma variável aleatória. O valor esperado de  $y$  para cada valor de  $x$  é [38]

$$E(y | x) = \beta_0 + \beta_1 x \quad (2.10)$$

Onde a intercepção  $\beta_0$  e o declive  $\beta_1$  são coeficientes de regressão desconhecidos. Assumimos que cada observação,  $y$ , pode ser descrita pelo modelo

$$y = \beta_0 + \beta_1 x + \epsilon \quad (2.11)$$

Onde  $\epsilon$  é um erro aleatório com média zero e variância (desconhecida)  $\sigma^2$ . Os erros aleatórios correspondentes a diferentes observações também são assumidos como variáveis aleatórias não correlacionadas.

As estimativas de  $\beta_0$  e  $\beta_1$  devem resultar em uma linha que é (em algum sentido) um "melhor ajuste" para os dados.

A análise de regressão é aplicada para compreender a dependência estatística de uma variável em relação a outras variáveis. A técnica pode mostrar que proporção de variação entre variáveis se deve à variável dependente e que proporção se deve às variáveis independentes. A relação entre as variáveis pode ser ilustrada graficamente ou, na maioria dos casos, por via de uma equação matemática.



### 2.2.1 Conformidade do modelo de regressão

Ajustar um modelo de regressão requer vários pressupostos. A estimativa dos parâmetros do modelo requer a suposição de que os erros são variáveis aleatórias não correlacionadas com variável média zero e constante. Testes de hipóteses e estimativa de intervalo exigem que os erros sejam normalmente distribuídos. Além disso, assumimos que a ordem do modelo está correta; isto é, se encaixarmos um modelo de regressão linear simples, estamos assumindo que o fenômeno realmente se comporta de maneira linear ou de primeira ordem. O analista deve sempre considerar a validade desses pressupostos como duvidosos e realizar análises para examinar a conformidade do modelo que foi obtido. Nesta seção, discutimos métodos úteis a este respeito.

#### 2.2.1.1 Análise residual

Os resíduos de um modelo de regressão são

$$e_i = y_i - \hat{y}_i, i = 1, 2, \dots, n \quad (2.12)$$

onde  $y_i$  é uma observação real e  $\hat{y}_i$  é o valor ajustado correspondente do modelo de regressão [38]. A análise dos resíduos é frequentemente útil para verificar a hipótese de que os erros são distribuídos aproximadamente normais, com variação constante e para determinar se os termos adicionais no modelo seriam úteis.

Como uma verificação aproximada da normalidade, o analista pode construir um histograma de frequência dos resíduos ou uma parcela de probabilidade normal de resíduos. Muitos programas de computador produzirão um gráfico de probabilidade normal de resíduos, e como os tamanhos de amostra em regressão são muitas vezes muito pequenos para que seja significativo para um histograma, o método de traçado de probabilidade normal é escolhido.

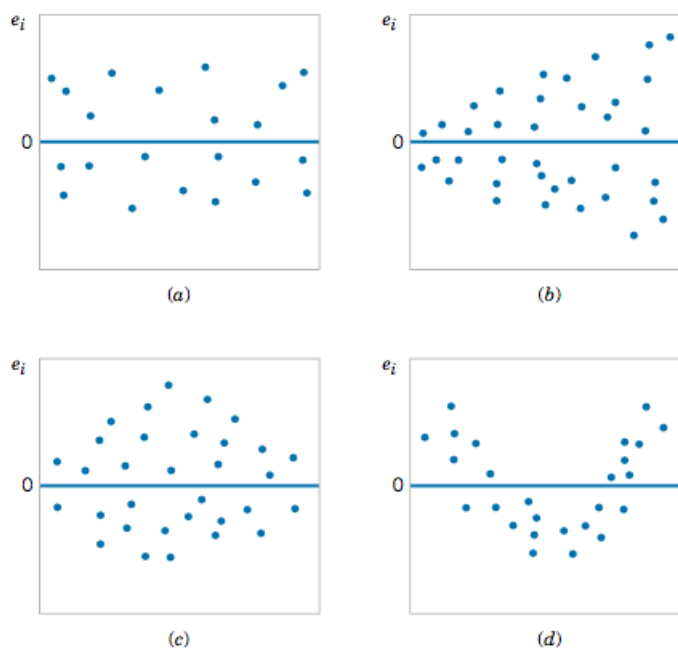


Figura 2.1: Gráficos de padrões residuais, (a) padrão uniforme e satisfatório e outros padrões não uniformes similares a (b) um funil, (c) em arco duplo e (d) flutuações sistemáticas de sinal.

É muito útil traçar os resíduos na seguinte ordem; (1) na sequência de tempo (se conhecida), (2) em função de  $\hat{y}$  e (3) contra a variável independente  $x$ . Esses gráficos geralmente parecerão com um dos quatro padrões gerais mostrados na figura 2.1. Padrão (a) na figura 2.1 é típico de um modelo bem ajustado, os padrões (b) e (c) são típicos de variações de aleatoriedade e (d) indicativo de que o modelo é inadequado. Se os resíduos aparecerem como em (b), a variação das observações pode aumentar com o tempo ou com a magnitude de  $y_i$  ou  $x_i$ . A transformação de dados na resposta  $y$  é frequentemente usada para eliminar esse problema. As transformações estabilizadoras de variância amplamente utilizadas incluem o uso de  $\sqrt{y}$ ,  $\ln y$ , ou  $1/y$  como resposta. Se um gráfico dos resíduos em função do tempo tiver a aparência de (b), a variação das observações está aumentando com o tempo. Parcelas de resíduos em função de  $\hat{y}_i$  e  $x_i$  que se parecem com (c) também indicam desigualdade de variância. Parcelas residuais que parecem com (d) indicam inadequação do modelo; ou seja, os termos de ordem superior devem ser adicionados ao modelo, uma transformação na variável  $x$  ou a variável  $y$  (ou ambas) deve ser considerada, ou outros regressores devem ser considerados.

### 2.3 Modelos de regressão polinomial

Mesmo sem saber a forma funcional do processo gerador de dados, podemos desenvolver uma técnica geral que funciona bem para problemas de relações

não lineares. Em primeiro lugar, é preciso lembrar que, teoricamente, podemos aproximar qualquer função com um polinómio. Então, podemos utilizar esse facto para estender regressão linear para regressão polinomial. A ideia é bastante simples: com um conjunto de variáveis existentes, construímos novas variáveis polinomiais e a regressão com elas terá mais capacidade quanto maior o grau do polinómio criado. É importante ressaltar que o algoritmo de regressão linear não muda! A única coisa que muda é que acrescentamos variáveis polinomiais, por exemplo, elevando o conjunto inicial de variáveis ao quadrado, ao cubo, à quarta, etc.

O modelo linear [39]

$$Y = X\beta + \epsilon \quad (2.13)$$

é um modelo geral que pode ser usado para ajustar qualquer relação linear nos parâmetros desconhecidos, neste caso o "X" representa a matriz de predictores enquanto o "Y" a matriz de respostas. Isso inclui a classe importante de modelos de regressão polinomial. Por exemplo, o polinómio de segundo grau em uma variável

$$Y = \beta_0 + \beta_1 x + \beta_{11} x^2 + \epsilon \quad (2.14)$$

E o polinómio de segundo grau em duas variáveis

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \epsilon \quad (2.15)$$

São modelos de regressão linear. Os modelos de regressão polinomial são amplamente utilizados quando a resposta é curvilínea, porque os princípios gerais de regressão múltipla podem ser aplicados. O termo  $\beta_{12}$  representa a interacção directa entre  $x_1$  e  $x_2$ .

Considerações finais. Uma grande limitação do modelo de regressão linear é que a máquina não aprende a forma da não linearidade, ficando essa parte sob responsabilidade da pessoa que está elaborar o modelo. Em outras palavras, é preciso codificar à força a não linearidade no algoritmo e isso é feito criando variáveis novas. Por algum tempo, essa foi a principal forma de aprender não linearidades, mas hoje existem métodos muito mais eficientes para isso. A segunda desvantagem do modelo de regressão polinomial é que ele não funciona bem quando temos muitas variáveis e o grau do polinómio é muito alto. Isso porque o número de variáveis cresce exponencialmente com o grau do polinómio ajustado, tornando os resultados rapidamente sem significado.

### 2.3.1 Seleção de variáveis e construção de modelos

Um problema importante em muitas aplicações da análise de regressão envolve a seleção do conjunto de variáveis de regressão a serem usadas no modelo. Em algumas situações, a experiência adquirida ou considerações teóricas subjacentes podem auxiliar o analista a especificar as variáveis estabelecidas ou regressoras a serem usadas em uma situação particular. No entanto, o problema consiste em selecionar um conjunto apropriado de regressores de um conjunto que provavelmente inclua todas as variáveis importantes, mas nem todos esses regressores candidatos são necessários para modelar adequadamente a resposta  $y$ .

Em tal situação, estamos interessados em selecionar as variáveis [39]; ou seja, rastreando as variáveis candidatas para obter um modelo de regressão que

contenha o "melhor" subconjunto de variáveis regressoras. Pretendemos que o modelo final contenha variáveis regressoras suficientes para que, no uso pretendido do modelo (previsão, por exemplo), ele funcionará de forma satisfatória. Por outro lado, para manter os custos de manutenção do modelo ao mínimo e para tornar o modelo fácil de usar, o modelo deve usar o menor número possível de variáveis regressores. O compromisso entre esses objetivos conflitantes geralmente é chamado de encontrar a "melhor" equação de regressão. No entanto, na maioria dos problemas, nenhum modelo de regressão simples é "melhor" em termos dos vários critérios de avaliação que foram propostos. Um grande julgamento e experiência com o sistema a ser modelado, geralmente é necessário para selecionar um conjunto apropriado de variáveis regressoras para uma equação de regressão.

Nenhum algoritmo produzirá sempre uma boa solução para o problema de seleção das variáveis. A maioria dos procedimentos atualmente disponíveis são técnicas de busca, e para realizar de forma satisfatória, eles precisam de interação com o analista [39]. Iremos discutir algumas das técnicas de seleção de variáveis mais populares. Se assumimos que existem  $K$  regressores de candidatos (parâmetros),  $x_1, x_2, \dots, x_k$  e uma única variável de resposta  $y$ . Todos os modelos incluirão um termo de intercepção  $\beta_0$ , de modo que o modelo com todas as variáveis incluídas terá termos  $K+1$ . Além disso, a forma funcional de cada variável candidata (por exemplo,  $x_1 = 1/x, x_2 = \ln x$ , etc.) é correta.

### 2.3.1.1 Todas as regressões possíveis

Esta abordagem exige que o analista se adapte a todas as equações de regressão envolvendo uma variável candidata, todas as equações de regressão envolvendo duas variáveis candidatas, e assim por diante. Então, essas equações são avaliadas de acordo com alguns critérios adequados para selecionar o "melhor" modelo de regressão. Se houver  $K$  regressores/parâmetros de candidatos, existem equações totais de  $2^K$  a serem examinadas. Por exemplo, se  $K=4$ , existem  $2^4=16$  possíveis equações de regressão; Enquanto que se  $K=10$ , existem  $2^{10}=1024$  equações de regressão possíveis. Assim, o número de equações a serem examinadas aumenta rapidamente à medida que o número de variáveis candidatas aumenta. No entanto, existem alguns algoritmos de computação muito eficientes para todas as possíveis regressões disponíveis e são amplamente implementados em software estatístico, por isso é um procedimento muito prático, a menos que o número de regressores de candidatos seja bastante grande.

Podem ser utilizados vários critérios para avaliar e comparar os diferentes modos de regressão obtidos.

### 2.3.1.2 Seleção de modelos por etapas

O procedimento constrói iterativamente uma sequência de modelos de regressão, adicionando ou removendo variáveis em cada etapa. O critério para adicionar ou remover uma variável em qualquer etapa geralmente é expresso em termos de um F-teste parcial. Considerando  $f_{in}$  o valor da variável F-aleatória para adicionar uma variável ao modelo e  $f_{out}$  o valor da variável F-aleatória para remover uma variável do modelo. Devemos ter  $f_{in} \geq f_{out}$ , e geralmente é  $f_{in} = f_{out}$ .

A regressão passo a passo começa por formar um modelo de uma variável

usando a variável regressor que possui a maior correlação com a variável de resposta  $y$ . Esse também será o regressor a estabelecer o maior F-teste. Por exemplo, suponha que nesta etapa,  $x_1$  esteja selecionado. No segundo passo, as variáveis candidatas K-1 restantes são examinadas e a variável para a qual a F-teste

$$F_j = \frac{SS_R(\beta_j|\beta_1, \beta_0)}{MS_E(x_j, x_1)} \quad (2.16)$$

É um máximo adicionado à equação, desde que  $f_j > f_{in}$ . Na equação 2.16,  $MS_E(x_j, x_1)$  refere o quadrado médio para o erro do modelo que contém ambas variáveis  $x_1$  e  $x_j$ . Suponha que este procedimento indique que o  $x_2$  deve ser adicionado ao modelo. Agora, o algoritmo de regressão por etapas determina se a variável  $x_1$  adicionada no primeiro passo deve ser removida. Isso é feito calculando o F-teste

$$F_1 = \frac{SS_R(\beta_1|\beta_2, \beta_0)}{MS_E(x_1, x_2)} \quad (2.17)$$

Se o valor calculado  $f_1 < f_{out}$ , a variável  $x_1$  é removida; Caso contrário, é mantida, e tentaremos adicionar um regressor ao modelo que contém ambas variáveis  $x_1$  e  $x_2$ .

Em geral, em cada etapa, o conjunto restante de candidatos regressores é examinado e o regressor com o maior F-teste é inserido, desde que o valor observado de  $f$  supere  $f_{in}$ . Em seguida, calcula-se F-teste para cada regressor no modelo, e o regressor com o menor valor observado de  $F$  é excluído se o  $f$  obtido  $< f_{out}$ . O procedimento continua até que nenhum outro regressor possa ser adicionado ou removido do modelo.

- (Forward Selection): Este procedimento de seleção é uma variação da regressão por etapas e baseia-se no princípio de que os regressores devem ser adicionados ao modelo um de cada vez até que não existam regressores candidatos permanentes que produzam um aumento significativo na soma de regressão dos quadrados. Ou seja, as variáveis são adicionadas uma de cada vez, desde que seu valor  $F$  supere o  $f_{in}$ . Esta seleção é uma simplificação da regressão por etapas que omite o F-teste para variáveis do modelo que foram adicionadas nas etapas anteriores. Esta é uma fraqueza do (Forward selection), ou seja, o procedimento não explora o efeito que a adição de um regressor no passo atual tem nas variáveis de regressor adicionadas em etapas anteriores.

-(Backward Elimination): O algoritmo deste procedimento de eliminação começa com todos K regressores candidatos no modelo. Em seguida, o regressor com a menor F-teste é excluído se este F-teste for insignificante, ou seja, se  $f < f_{out}$ . Em seguida, o modelo com K-1 regressores está pronto, e o próximo potencial regressor para eliminação será encontrado. O algoritmo termina quando nenhum outro regressor pode ser excluído.

### 2.3.2 Critérios

Nesta seção irei, abordar alguns dos critérios que podem ser utilizados para o processo de modelar adequadamente as variáveis de resposta de um conjunto de regressores/preditores. Estes são, o método dos mínimos quadrados, permite um melhor ajuste para um conjunto de dados tentando minimizar a

soma dos quadrados das diferenças entre o valor estimado e os dados observados (tais diferenças são chamadas resíduos) permite minimizar a soma dos quadrados dos resíduos, o coeficiente de determinação ( $R^2$ ), permite observar a resposta de um modelo se adequa ao conjunto de regressores, o AIC é baseada na força da evidência e fornece a cada um dos modelos candidatos uma medida de incerteza e por último o BIC, também tem os mesmos pressupostos que AIC mas com uma diferença o termo de penalidade é maior neste caso.

### 2.3.2.1 Mínimos Quadrados

Os métodos assim chamados partem do mínimo da soma  $S$ , dos quadrados das distancias verticais entre os pontos experimentais e a reta de regressão [40]:

$$S = \sum w_i (y_i - \alpha - \beta x_i)^2 \quad (2.18)$$

Aqui os ponderadores  $w_i$  são utilizados para que seja possível diferenciar a influência de um determinado ponto sobre as estimativas dos parâmetros. Para as variáveis encontradas na aplicação deste método, estes pesos recebem atribuições diferentes. Normalmente, são utilizados pesos que se relacionam com a precisão das medidas ( $x_i, y_i$ ), através de suas variâncias  $\alpha_{x_i}^2$  e  $\alpha_{y_i}^2$ .

Para se obter as estimativas para  $\alpha$  e  $\beta$  que minimizam (2.18), escrevem-se as derivadas parciais de  $S$  em relação aos dois parâmetros e estas são igualadas a zero. Estas são as chamadas equações normais de regressão, e o sistema formado por elas deve ser resolvido para se obter os estimadores desejados. Dependendo da definição dos pesos  $w_i$  acima, este sistema pode ser mais imediatamente resolvido ou não.

### 2.3.2.2 Coeficiente de determinação ( $R^2$ )

O coeficiente de determinação é frequentemente usado para avaliar a adequação de um modelo de regressão [41]. Posteriormente, veremos que no caso em que  $x$  e  $y$  são variáveis aleatórias distribuídas conjuntamente,  $R^2$  é o quadrado do coeficiente de correlação entre  $x$  e  $y$ .

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \quad (2.19)$$

A partir da análise da variância da identidade na equação 2.19,  $0 \leq R^2 \leq 1$ . Muitas vezes nos referimos vagamente a  $R^2$  como a quantidade de variabilidade nos dados explicados ou explicados pelo modelo de regressão.

A estatística  $R^2$  deve ser usada com cautela, pois é possível aumentar a unidade  $R^2$  simplesmente adicionando termos suficientes ao modelo. Por exemplo, podemos obter um ajuste "perfeito" para  $n$  pontos de dados com um polinômio de grau  $n-1$ . Além disso, o  $R^2$  sempre aumentará se adicionarmos uma variável ao modelo, mas isso não implica necessariamente que o novo modelo seja superior ao antigo. A menos que a soma de erro dos quadrados no novo modelo seja reduzida em um valor igual ao quadrado médio do erro original, o novo modelo terá um erro maior, o quadrado médio do que o antigo, devido à perda de um grau de liberdade de erro. Assim, o novo modelo será realmente pior do que o antigo.

Existem vários equívocos sobre o  $R^2$ . Em geral,  $R^2$  não mede a magnitude da inclinação da linha de regressão. Um grande valor de  $R^2$  não implica uma

inclinação íngreme. Mesmo que  $y$  e  $x$  estejam relacionados de forma não-linear,  $R^2$  geralmente será grande. Finalmente, embora  $R^2$  seja grande, isso não implica necessariamente que o modelo de regressão forneça previsões precisas de futuras observações.

### 2.3.2.3 AIC

O critério de informação Akaike (AIC) foi desenvolvido pela primeira vez por Akaike [42] como uma forma de comparar diferentes modelos em um determinado resultado. Por exemplo, se prendermos saber quais as variáveis que influenciam determinados modelos de regressão. Muitas vezes, porém, o próprio modelo é tratado como apenas uma ferramenta para revelar a relação entre o resultado e uma variável específica. A seleção do modelo é importante, uma vez que o modelo subjacente não pode capturar a verdadeira natureza da variabilidade na variável de resultado, enquanto um modelo superado perde a generalidade. AIC é então uma maneira de selecionar o modelo que melhor equilibra essas desvantagens. Uma vez que um melhor modelo é selecionado, o teste de hipóteses nulas tradicionais pode ser usado no melhor modelo para determinar a relação entre variáveis específicas e o resultado de interesse.

Akaike mostrou que esta seleção do “melhor” modelo é determinado por uma pontuação AIC [43]:

$$AIC = 2p - 2 \ln(\hat{L}) \quad (2.20)$$

Onde  $p$  é o número de parâmetros estimáveis (graus de liberdade) e  $(\hat{L})$  é função de máxima verossimilhança do modelo estimado. O "2" constante permanece "por razões históricas" [44].[45] refinaram ainda mais esta estimativa para corrigir pequenas amostras de dados [43]:

$$AIC_c = AIC + \frac{2p(p+1)}{n-p-1} \quad (2.21)$$

Onde  $n$  é o tamanho da amostra e  $p$  e AIC são definidos acima. Se  $n$  for grande em relação a  $p$ , esta correção é insignificante e AIC é suficiente. AICc é mais abrangente, no entanto, é geralmente usado no lugar do AIC. O melhor modelo é então o modelo com a pontuação AICc (ou AIC) mais baixa. É importante notar que os valores de AIC e AICc são ordinais, indicam uma hierarquia numa determinada sequência e não significam nada por conta própria. Eles são simplesmente uma maneira de classificar os modelos.

Embora a AIC seja uma ferramenta poderosa para a comparação de modelos, permaneceu bastante desconhecida no ocidente durante muitos anos, devido ao fato de que a pesquisa original e muitos artigos sobre a construção foram escritos em japonês e não foram traduzidos para o inglês.[44] forneceram o primeiro exame completo da AIC e dos critérios de informação em geral e seu texto é amplamente considerado como uma autoridade sobre o assunto. O AIC é amplamente utilizado nas ciências biológicas, especificamente nas ciências ambientais e marinhas e nas bacias hidrográficas. Em contrapartida, a seleção do modelo que aplica AIC é baseada na força da evidência e fornece a cada um dos modelos candidatos uma medida de incerteza. O que é importante enfatizar é que a AIC pode identificar qual modelo é o melhor entre os modelos candidatos, mas isso não significa que nenhum dos modelos candidatos faça um bom trabalho para explicar os dados.

#### 2.3.2.4 BIC

O critério de informação bayesiano (BIC), proposto por Schwarz [46] e, portanto, é também referido como o critério de informação Schwarz e critério de informação Schwarz Bayesiano.

Este Critério de informação, é outro critério de seleção de modelo baseado na teoria da informação, mas definido dentro de um contexto bayesiano. Em estatística, o critério de informação Bayesiano ou critério Schwarz é um critério para seleção de modelos entre um conjunto finito de modelos. Baseia-se, em parte, na função de verossimilhança, e está intimamente relacionado ao critério de informação de AIC. Ao construir modelos, é possível aumentar a probabilidade adicionando parâmetros, mas, assim, pode resultar em superposição. O BIC resolve este problema introduzindo um termo de penalidade para o número de parâmetros no modelo. O termo de penalidade é maior no BIC do que no AIC.[43] fornecem argumentos teóricos a favor do AIC, particularmente o AICc sobre o BIC. Além disso, no caso da análise de regressão multivariada, [47] explica por que a AIC é melhor do que o BIC na seleção do modelo.

O BIC é calculado da seguinte forma [43]

$$BIC = p \ln(n) - 2 \ln(\hat{L}) \quad (2.22)$$

Onde os termos acima são os mesmos descritos que na descrição da AIC. O melhor modelo é aquele que fornece o BIC mínimo.



## Capítulo 3

# Experimental

Neste capítulo explica-se em detalhe todos os procedimentos envolvidos, no tratamento das variáveis em estudo, foram utilizadas como ferramentas de apoio, o R-studio, excel e numbers. Com este conjunto ferramentas e dados podemos assim alcançar o objectivo proposto do trabalho, estabelecer padrões de emissões de poluentes através de modelos.

### 3.1 Sistematização de processos

O processo de construção para posterior tratamento foi conduzido, através da elaboração de uma matriz inicial composta pelos parâmetros de qualidade de ar, os dados de medições por um período correspondente de 28 dias, referentes ao mês de fevereiro de 2017 recolhidas ao longo de cada hora no mesmo período, perfazendo um ciclo diário, em seguida, de modo a facilitar o mecanismo de tratamento e processamento de dados, cada variável foi codificada através da atribuição de uma letra, ver tabela 3.1

Tabela 3.1: Correspondência de termos para o conjunto dos compostos da qualidade de ar em análise.

|    |                 |                |                   |                  |                 |    |                 |       |      |
|----|-----------------|----------------|-------------------|------------------|-----------------|----|-----------------|-------|------|
| CO | NO <sub>2</sub> | O <sub>3</sub> | PM <sub>2.5</sub> | PM <sub>10</sub> | SO <sub>2</sub> | NO | NH <sub>3</sub> | NMVOC | PANs |
| a  | b               | c              | d                 | e                | f               | g  | h               | i     | j    |

Após esta transformação, de modo a aumentar a fiabilidade do método, foram adicionados na matriz termos de raiz quadrada ver tabela 3.2 e de potência 2 ver tabela 3.2 para cada um dos elementos.

Tabela 3.2: Nomenclatura para as dependências dos respectivos dados dos compostos da qualidade de ar.

| Dependências           |                                    |        |    |    |    |    |    |    |    |    |    |
|------------------------|------------------------------------|--------|----|----|----|----|----|----|----|----|----|
| em 1 <sup>o</sup> grau | 10                                 | a      | b  | c  | d  | e  | f  | g  | h  | i  | j  |
| em 2 <sup>o</sup> grau | 10                                 | aa     | bb | cc | dd | ee | ff | gg | hh | ii | jj |
| em grau 1/2            | 10                                 | ra     | rb | rc | rd | re | rf | rg | rh | ri | rj |
| em grau 0              | 1                                  | Intcp. |    |    |    |    |    |    |    |    |    |
| Interacções            | 45                                 | ab     | ac | ad | ae | af | ag | ah | ai | aj | bc |
|                        |                                    | bd     | be | bf | bg | bh | bi | bj | cd | ce | cf |
|                        |                                    | cg     | ch | ci | cj | de | df | dg | dh | di | dj |
|                        |                                    | ef     | eg | eh | ei | ej | fg | fh | fi | fj | gh |
|                        |                                    | gi     | gj | hi | hj | ij |    |    |    |    |    |
| Parâmetros             | 76(total de estimativas do modelo) |        |    |    |    |    |    |    |    |    |    |

De modo a permitir uma melhor interpretação a nível significativo no que se refere à variação da resposta (predita) foram criados os termos de interacção ver tabela 3.2.

Os termos de interacção refletem um efeito diferencial na resposta advindo da relação entre duas variáveis, mesmo que a inclusão da interacção "anule" a significância dos efeitos principais. Este processo, bem como os anteriores, visam aumentar a variedade de aplicações que um mesmo modelo estatístico pode ter.

## 3.2 Organização da informação

Com este conjunto de operações obtemos um total, perfazendo com o numero inicial de termos, 76 regressores/preditores, este valor permite um melhor controle no ajuste do modelo através do incremento de população.

Na etapa seguinte, a informação relativa aos parâmetros de humidade relativa, temperatura, velocidade do vento, direcção do vento e precipitação, recolhida de uma sonda meteorológica, fonte in-situ, foi pre-tratada, antes de ser adicionada aos restantes conjunto de regressores, através do software Numbers, os dados da fonte estavam elaborados para medições recolhidas a cada 15min ao longo do dia, nesta fase pretendemos que estejam em medições de medias horárias, para assim estarem de acordo com estabelecido e correspondentes dados de parâmetros de qualidade de ar.

## 3.3 Pré-tratamento das variáveis

Possuímos agora uma matriz com toda a informação relevante agrupada, 81 regressores/preditores, no entanto precisamos de encontrar uma forma de obter um sistema equivalente para todo o conjunto de variáveis, para isso é necessário um escalamento.

Por vezes o efeito de escala faz-se sentir de tal forma que o refinamento do modelo pode ser comprometido. Este tipo de anomalia pode surgir na modelação quadrática dado que quando algumas das variáveis já apresentam valores um pouco acima do normal, o seu quadrado pode fazer deslocar esses valores para ordens de grandeza superiores e, como tal, a sua influência nos resíduos

do modelo irá ser afectada. Para colocar todas as variáveis na mesma escala procedeu-se ao seu escalamento através de:

$$Z_i = \frac{x_i - x_{0i}}{x_{1i} - x_{0i}} \quad (3.1)$$

onde  $x_{0i}$  e  $x_{1i}$  correspondem aos extremos inferior e superior da variável  $x_i$ .

### 3.4 Modelação

Podemos agora partir para a fase de modelação para cada um dos parâmetros de qualidade de ar, o problema da identificação do modelo é escolher entre um conjunto de modelos candidatos, um para descrever um determinado conjunto de dados, a estratégia aplicada será através da pesquisa descendente por etapas (Forward e Backward) com os critérios AIC e BIC partindo de um modelo completo para um modelo simples, ou seja, construindo um modelo em que apresenta todo o conjunto de preditores removendo as variáveis que pouco contribui e interferem no processo de resposta(predita), tais como os termos de interacção, termos de ordem 2 e raiz quadrada em que participa o parâmetro que pretendemos modelar na análise, não deve ficar no modelo inicial.

A estes novos modelos, após análise efetuada no R-studio do impacto de cada preditor através dos critérios referidos, podemos seleccionar o melhor modelo [48].

# Capítulo 4

## Resultados

O estudo em causa foi realizado sobre dados ambientais e da qualidade do ar fornecidos através do cruzamento de informação obtida pelo Copernicus e a sonda in-situ (projecto SOUL) no mês de fevereiro de 2017. Os resultados obtidos referem-se a valores médios horários diários sendo o número de pontos para cada variável 672 referente a 24 resultados diários registados ao longo de 28 dias.

### 4.1 Diagnóstico preliminar

Numa fase inicial pretende-se caracterizar cada variável em termos de variabilidade no período de observação e respectivo ciclo diário.

Na tabela 4.1 encontram-se listadas algumas características e estimativas relacionadas com as variáveis ambientais - temperatura, humidade, precipitação, velocidade do vento, etc. Na tabela 4.2 encontram-se registadas as variáveis relacionadas com o diagnóstico da qualidade do ar obtidas no mesmo período.

Tabela 4.1: Listagem das variáveis utilizadas na caracterização do ambiente durante o mês de fevereiro de 2017.

| código      | Temp        | Hrel              | Vvento              | Dvento            | Precip       |
|-------------|-------------|-------------------|---------------------|-------------------|--------------|
| variável    | temperatura | humidade relativa | velocidade do vento | direcção do vento | precipitação |
| unidades    | °C          | –                 | $m.s^{-1}$          | °                 | $mL.m^{-2}$  |
| máximo      | 22.9        | 100.0             | 14.5                | 359.8             | 7.4          |
| $X_{0.975}$ | 20.1        | 100.0             | 11.7                | 353.6             | 1.4          |
| $X_{0.500}$ | 12.1        | 87.8              | 2.6                 | 133.9             | 0.0          |
| $X_{0.025}$ | 5.9         | 44.1              | 0.0                 | 0.0               | 0.0          |
| mínimo      | 4.4         | 29.2              | 0.0                 | 0.0               | 0.0          |
| média       | 12.4        | 82.2              | 3.4                 | 155.1             | 0.1          |
| assimetria  | 0.3         | -0.9              | 1.2                 | 0.2               | 7.3          |
| curtose     | 3.5         | 2.8               | 4.3                 | 1.9               | 69.7         |

$X_a$  - percentil da distribuição

Tabela 4.2: Listagem das variáveis utilizadas no diagnóstico da qualidade do ar no mês de fevereiro de 2017.

| código      | a               | b               | c               | d                | e                |
|-------------|-----------------|-----------------|-----------------|------------------|------------------|
| variável    | CO              | NO <sub>2</sub> | O <sub>3</sub>  | PM <sub>25</sub> | PM <sub>10</sub> |
| unidades    | $\mu g.m^{-3}$  | $\mu g.m^{-3}$  | $\mu g.m^{-3}$  | $\mu g.m^{-3}$   | $\mu g.m^{-3}$   |
| máximo      | 271.7           | 31.2            | 100.1           | 36.6             | 45.4             |
| $X_{0.975}$ | 244.7           | 20.9            | 95.9            | 24.0             | 36.8             |
| $X_{0.500}$ | 164.3           | 4.8             | 76.0            | 6.7              | 11.7             |
| $X_{0.025}$ | 131.0           | 1.4             | 37.7            | 2.6              | 3.8              |
| mínimo      | 108.3           | 1.0             | 30.0            | 2.0              | 2.9              |
| média       | 169.2           | 6.3             | 73.3            | 7.9              | 13.5             |
| assimetria  | 1.33            | 1.9             | -0.7            | 2.4              | 1.6              |
| curtose     | 5.6             | 7.4             | 3.0             | 9.9              | 6.0              |
| variável    | f               | g               | h               | i                | j                |
| unidades    | SO <sub>2</sub> | NO              | NH <sub>3</sub> | NMVOC            | PANS             |
| máximo      | $\mu g.m^{-3}$  | $\mu g.m^{-3}$  | $\mu g.m^{-3}$  | $\mu g.m^{-3}$   | $\mu g.m^{-3}$   |
| $X_{0.975}$ | 4.5             | 5.7             | 1.9             | 34.0             | 2.7              |
| $X_{0.500}$ | 3.5             | 2.2             | 1.6             | 26.5             | 2.3              |
| $X_{0.025}$ | 1.4             | 0.2             | 0.4             | 8.2              | 1.0              |
| mínimo      | 0.5             | 0.0             | 0.0             | 3.6              | 0.4              |
| média       | 0.3             | 0.0             | 0.0             | 2.7              | 0.2              |
| assimetria  | 1.6             | 0.4             | 0.6             | 9.6              | 1.1              |
| curtose     | 0.8             | 3.4             | 0.9             | 1.8              | 1.0              |

$X_a$  - percentil da distribuição

O tipo de distribuição de valores é um prenúncio de como essa variável se comporta no tempo - variáveis platicúrticas ou de iso-probabilidade apresentam um conjunto de valores de similar probabilidade; já variáveis leptocúrticas estão relacionadas com variáveis em que existe uma muito elevada probabilidade de obter uma pequena gama de valores dentro da larga janela de valores possíveis; uma variável normal tende a apresentar com maior frequência um determinado valor central sendo também prováveis valores acima e a baixo deste valor de maior probabilidade.

A avaliar pelas estimativas obtidas na tabela 4.1 as variáveis ambientais temperatura (Temp) e humidade relativa (Hrel) apresentam-se como sendo distribuições normais enquanto que a direcção do vento (Dvento) tem uma nítida tendência para possuir uma distribuição platicúrtica e a variável precipitação (Precip) características de distribuição leptocúrtica.

Já no que respeita às variáveis de diagnóstico da qualidade do ar, tabela 4.2, as variáveis ozono (c), dióxido de enxofre (f), amoníaco (h) e PANS (j) apresentam uma tendência para a normalidade e as restantes uma tendência para a distribuições mais afiladas.

Numa primeira fase de diagnóstico inicial avalia-se a variação total e diária de cada uma das variáveis ambientais (a-j).

De seguida vamos analisar, caso a caso, cada uma destas dez variáveis de diagnóstico da qualidade do ar.

### 4.1.1 Monóxido de carbono

A variável monóxido de carbono (CO, codificada como “a”) registada ao longo do período de tempo considerado encontra-se representada na figura (4.1).

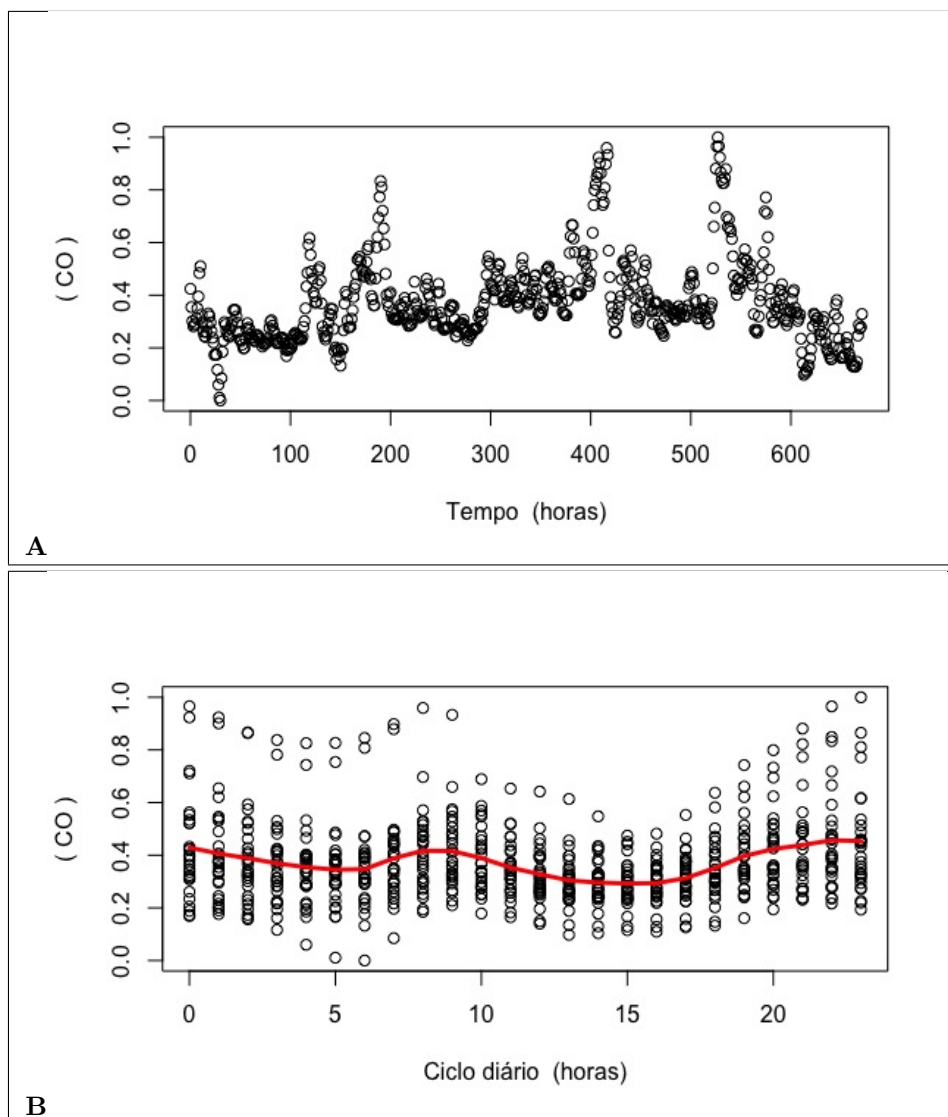


Figura 4.1: Representação dos valores relativos de monóxido de carbono obtidos (A) durante o período de tempo considerado (em horas) e (B) segundo a periodicidade diária (escala diária de 0-23h).

Da figura 4.1A) verifica-se que há uma tendência para um termo aproximadamente constante sobre o qual surgem alguns máximos e, em alguns casos, mínimos. Já segundo uma perspectiva de ciclo diário, figura 4.1B), observa-se uma tendência para maiores níveis de CO durante o período noturno (19h-2h) e um ciclo adicional das 7h-10h.

O ciclo menos intenso (7h-10h) pode estar relacionado com o aumento de tráfego rodoviário [49] e reinício da actividade laboral. Já o ciclo nocturno mais pronunciado (19h-2h) pode estar relacionado com a ausência de fotossíntese (incapacidade de consumo de dióxido de carbono), o regresso a casa (final do dia de trabalho e activação de sistemas de aquecimento) e alguma permanência dos níveis em indústrias que laboram até mais tarde.

O mínimo nos níveis de monóxido de carbono surgem por volta das 5h-6h quando a actividade Humana é mínima e começa a haver alguma claridade para reiniciar o processo de captura de dióxido de carbono através da fotossíntese.

Das figuras 4.1A) e 4.1B) verifica-se ainda que no mês de fevereiro (28 dias) existem pelo menos 4 dias onde os valores são extremos com tendência para atingir valores muito acima do normal enquanto que também existem pelo menos 2 dias com valores mínimos. Estas observações são também consistentes com uma distribuição leptocúrtica, caracterizada por uma predominância de valores centrais.

#### 4.1.2 Dióxido de azoto

O dióxido de azoto ( $NO_2$ , variável codificada como “b”) também apresenta uma variabilidade temporal notória, ver figura 4.2.

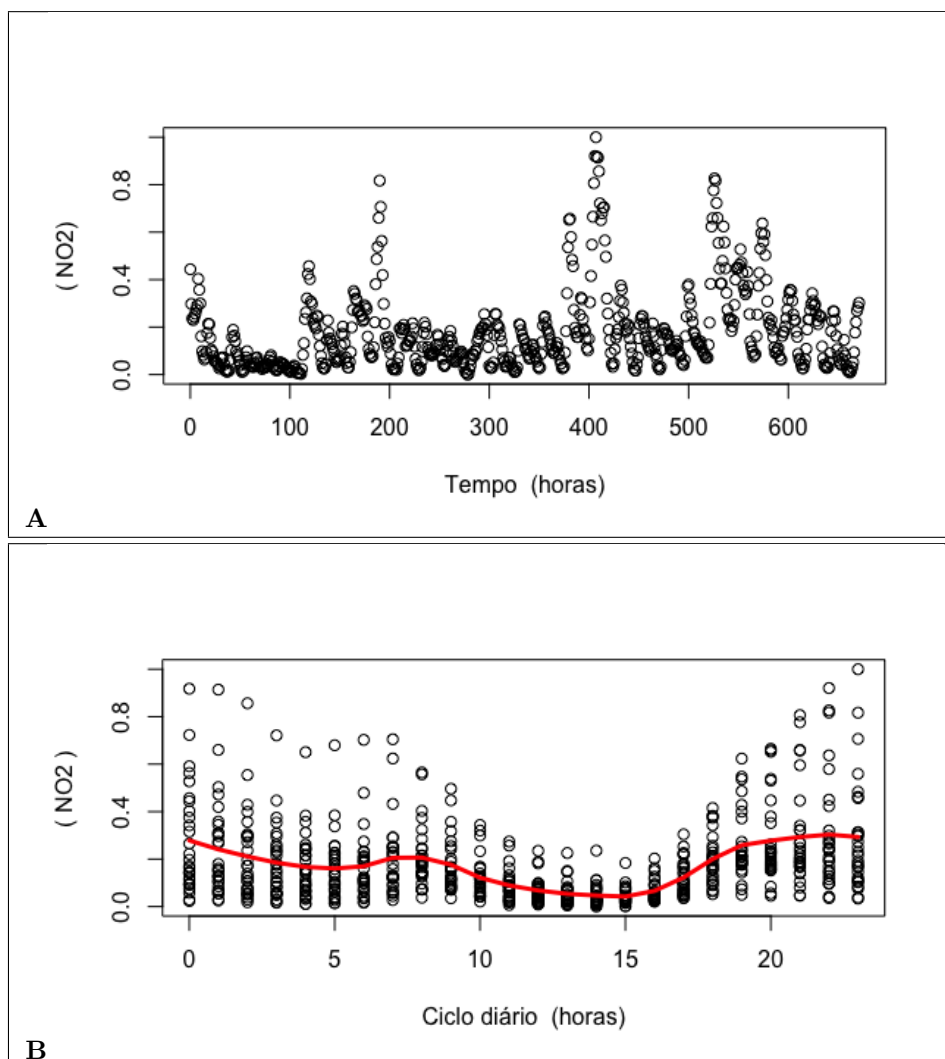


Figura 4.2: Representação dos valores relativos ao teor de dióxido de azoto obtidos (A) durante o período de tempo considerado (em horas) e (B) segundo a periodicidade diária (escala diária de 0-23h).

Da figura 4.2 observa-se que além de algumas discrepâncias em termos de valores máximos recordes, há também uma tendência cíclica diária, ver 4.2B) - uma forma sinusoidal suave com máximo por volta das 21-22h e o mínimo por volta das 14h; uma protuberância mais suave com máximo por volta das 8h. Provavelmente este segundo ciclo estará relacionado com o início da actividade Humana (transportes pesados) [50].

Este tipo de variabilidade está relacionada com uma distribuição tipicamente leptocúrtica caracterizada por um valor tendencialmente constante sobre o qual são sobrepostas contribuições que promovem a assimetria positiva, ver tabela 4.2.



### 4.1.3 Ozono

Os valores relativos do teor atmosférico em ozono ( $O_3$ , variável codificada através de “c”) encontra-se representado na figura 4.2.

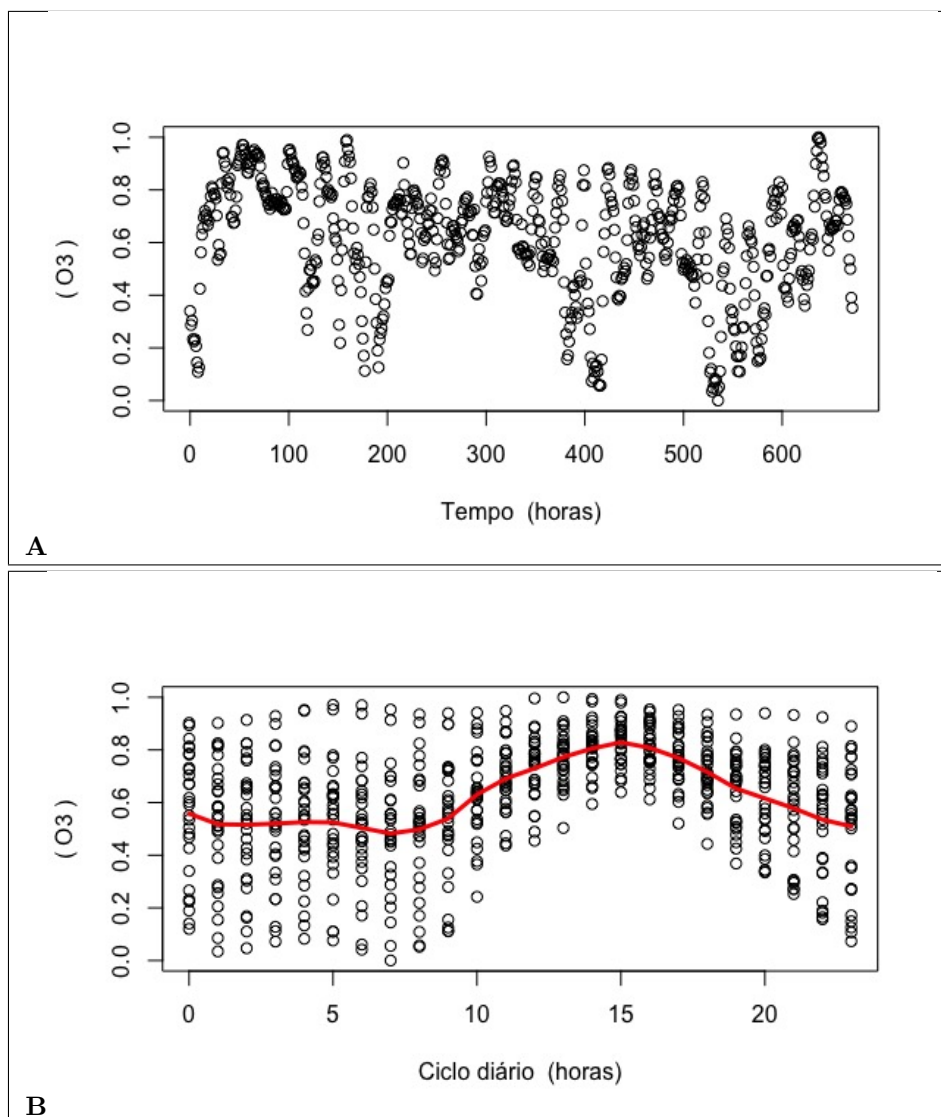


Figura 4.3: Representação dos valores relativos de ozono obtidos (A) durante o mês de fevereiro de 2017 (registado em horas) e (B) segundo a periodicidade diária (escala diária de 0-23h).

Da figura 4.3 verifica-se que existe uma enorme aleatoriedade de valores à excepção do período situado entre as 12-17h onde a tendência é muito mais coerente e consistente em valores mais elevados. A elevada aleatoriedade observada nesta variável é compatível com as características normais evidenciadas por esta distribuição, ver tabela 4.2.

Em termos de interpretação destes valores máximos consistentes apenas constatamos que tendem a coincidir com o período de máxima luminosidade em termos de índices médio de UV observada em Coimbra (11-16h) neste período [<https://www.ipma.pt/en/ambiente/uv/locais/index.jsp>]. Contudo, o conjunto de dados disponibilizado não possuía um registo consistente e fiável destes valores.

#### 4.1.4 $PM_{2.5}$

Os valores relativos do teor atmosférico de matéria particulada mais fina (de dimensão inferior a  $2.5 \mu\text{m}$ , variável codificada através de “**d**”) encontra-se representado na figura 4.4.

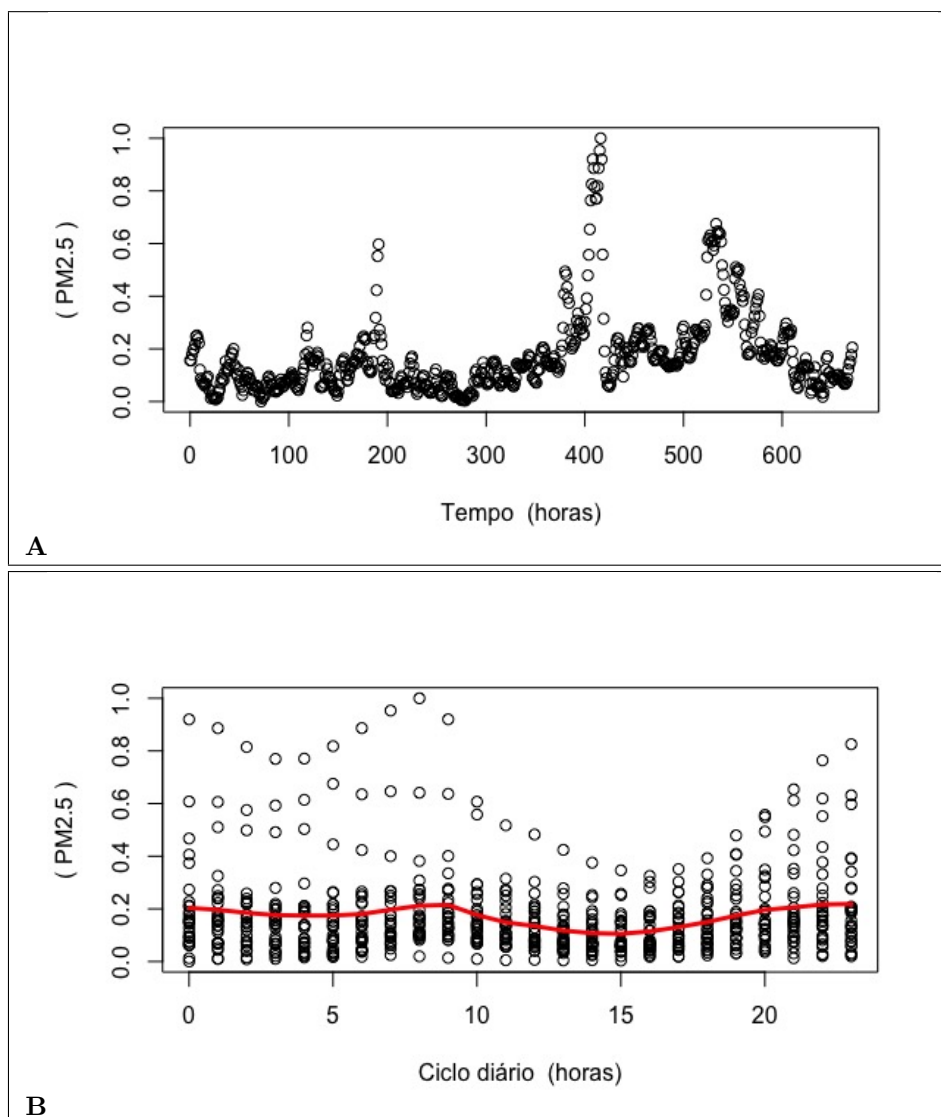


Figura 4.4: Representação dos valores relativos de PM2.5 obtidos (A) durante o período de tempo considerado (em horas) e (B) segundo a periodicidade diária (escala diária de 0-23h).

Desta representação evidencia-se uma tendência para valores com alguma constância estando presentes alguns períodos de repetição diária - tendência para um ligeiro mínimo por volta das 3-5h, valor máximo por volta das 8-9h, de novo um mínimo entre as 13-16h (período de maior intensidade luminosa) seguido de um aumento progressivo com máximo por volta das 21-2h.

Sobrepostos a este ciclo diário existem algumas situações extremas, patentes na figura 4.4A).

Devido à elevada consistência dos valores obtidos neste período temporal a distribuição destes valores apresentam-se também como sendo caracterizados com uma distribuição leptocúrtica com assimetria positiva reforçada pelas

anomalias em valores extremos altos.

#### 4.1.5 $PM_{10}$

Os valores relativos do teor atmosférico de matéria particulada menos fina ( $PM_{10}$ , variável codificada através de “e”) encontra-se representada na figura 4.5.

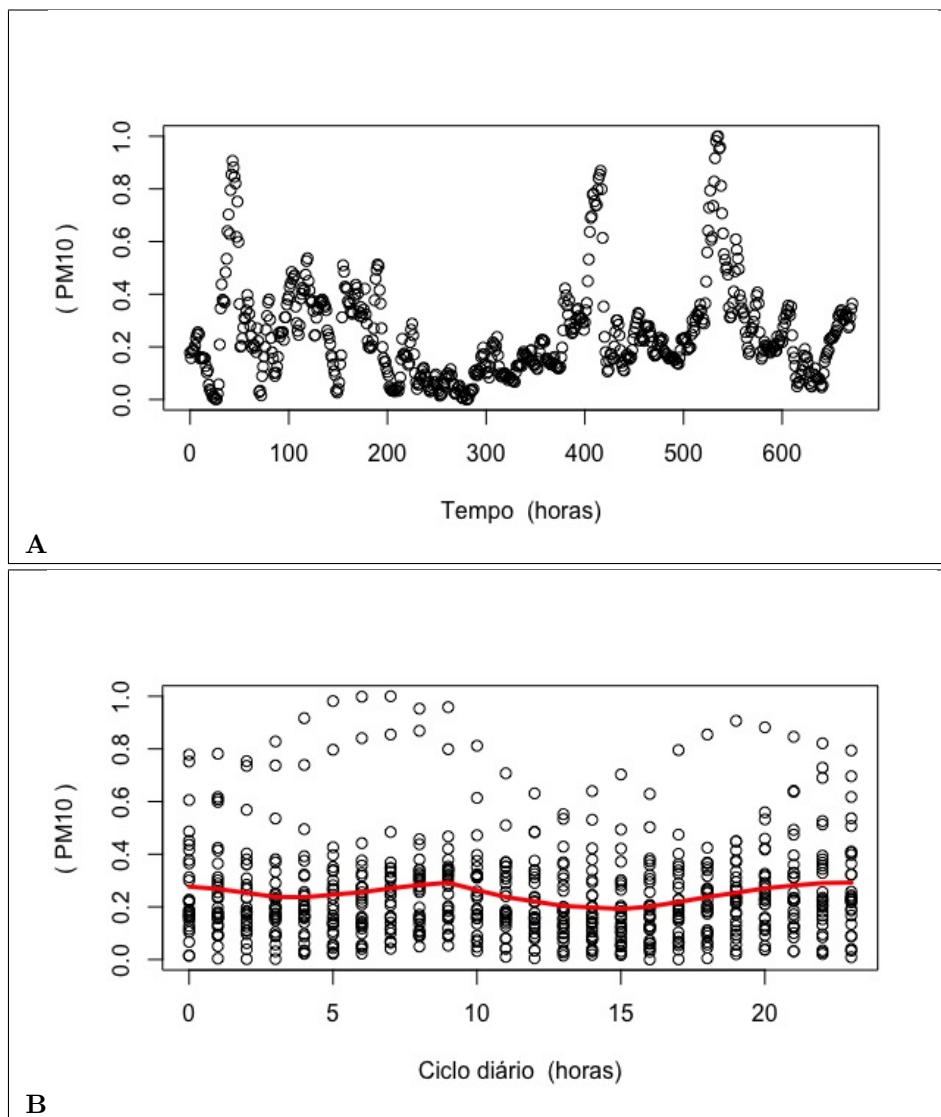


Figura 4.5: Representação dos valores relativos de  $PM_{10}$  obtidos (A) durante o período de tempo considerado (em horas) e (B) segundo a periodicidade diária (escala diária de 0-23h).

Apesar das anomalias observadas em curtos períodos de tempo (seqüências rápidas de valores máximos), existe uma tendência para que esta variável

apresente um comportamento tendencialmente constante e coerente com um determinado ciclo diário onde os valores máximos são atingidos nos períodos das 7-10h e 20-1h e os mínimos entre as 3-5h e 14-17h.

As características evidenciadas nesta representação são coerentes com uma distribuição também leptocúrtica com tendência para valores acima do máximo de probabilidade.

#### 4.1.6 Dióxido de enxofre

Os valores relativos do teor atmosférico em ozono ( $SO_2$ , variável codificada através de “f”) encontra-se representado na figura 4.6.

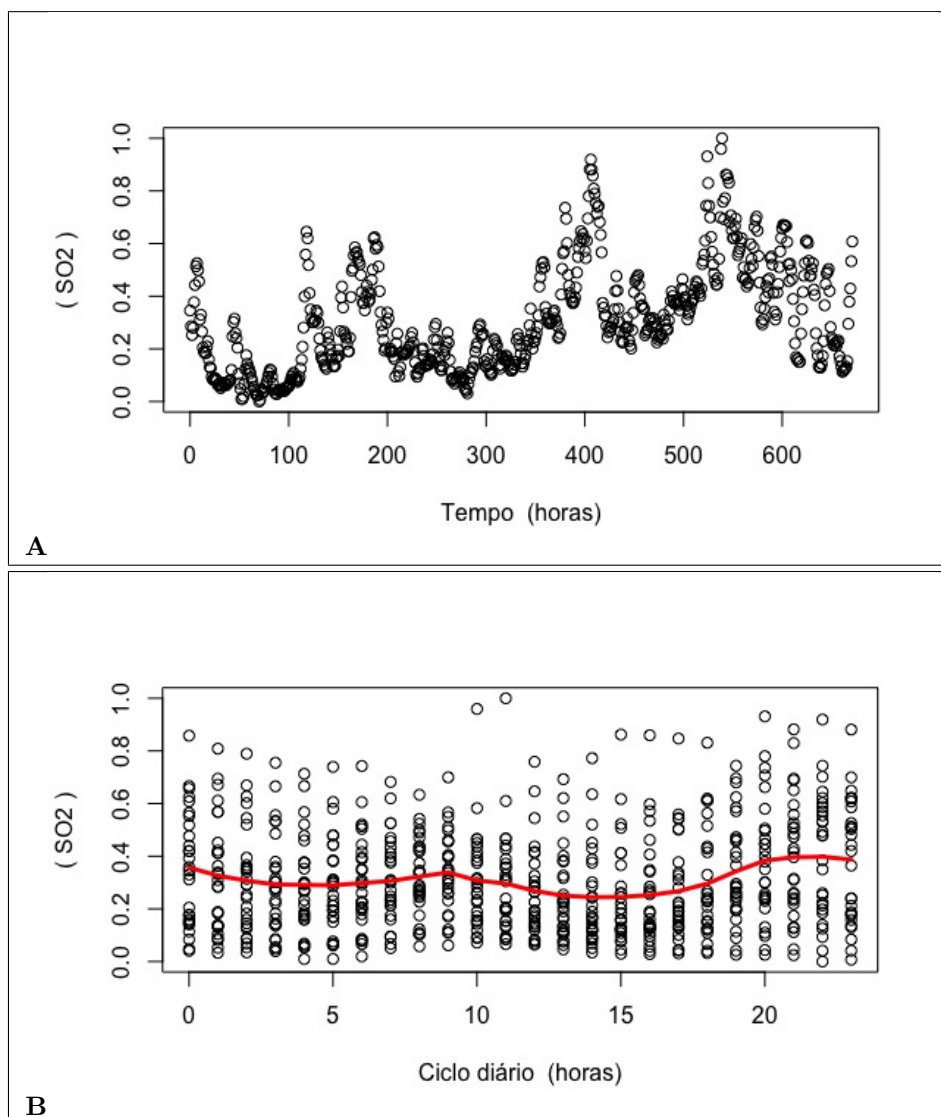


Figura 4.6: Representação dos valores relativos de dióxido de enxofre obtidos (A) durante o período de tempo considerado (em horas) e segundo a periodicidade diária (escala diária de 0-23h).

Durante o período de observação parece ter surgido quatro situações onde estes valores estiveram bem acima do valor normal, ver figura 4.6A). Já da figura 4.6B) verifica-se uma enorme aleatoriedade nos valores obtidos. Contudo, com base nos valores horários médios, linha vermelha sobreposta da figura 4.6B), parece também haver dois períodos onde a predominância de valores máximos ocorre - entre as 7-10h e 19-1h.

Devido à sua elevada estocasticidade, esta variável aproxima-se de uma distribuição normal.

### 4.1.7 Óxido de azoto

Os valores relativos do teor atmosférico em óxido de azoto (NO, variável codificada através de “g”) encontram-se representados na figura 4.7.

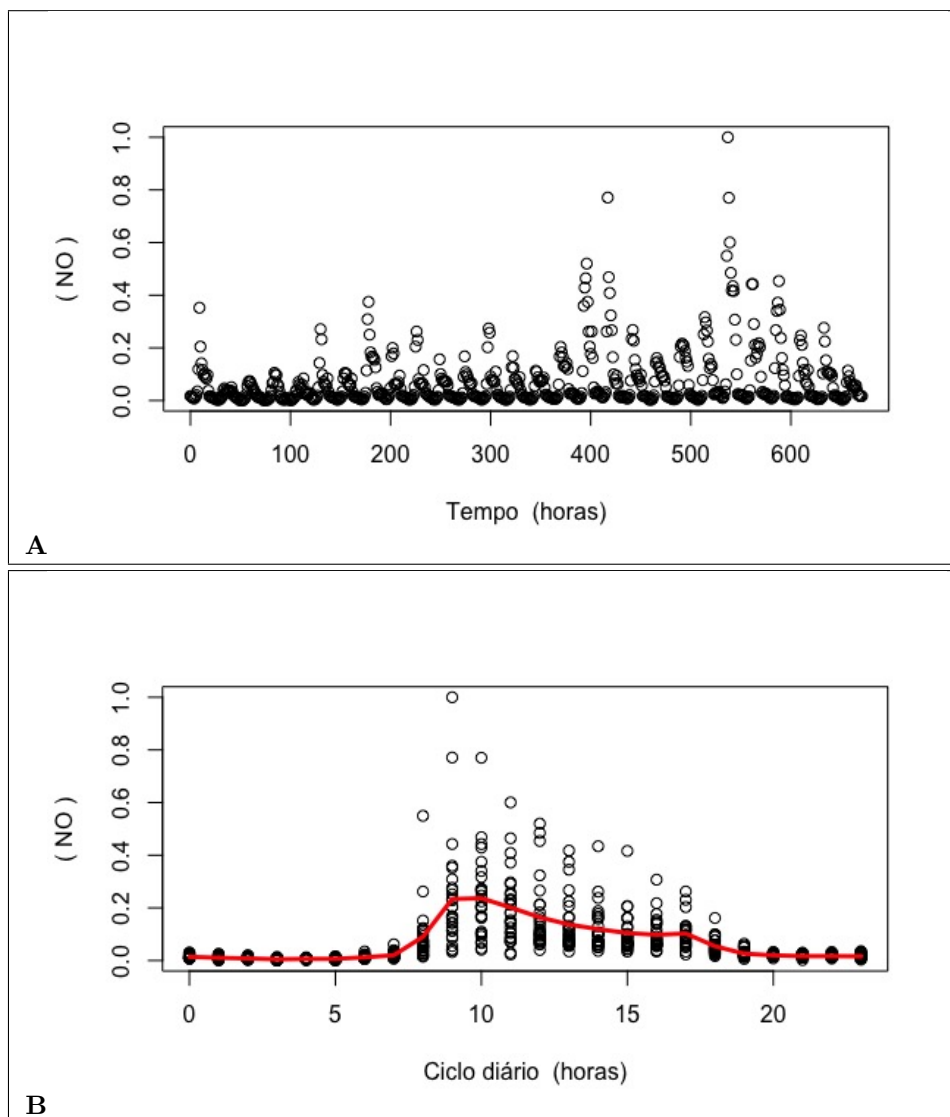


Figura 4.7: Representação dos valores relativos de monóxido de azoto obtidos (A) durante o período de tempo considerado (em horas) e (B) segundo a periodicidade diária (escala diária de 0-23h).

Os níveis de dióxido de azoto têm uma enorme consistência no período diário situado entre as 20-6h do dia onde atinge os valores mais baixos. Já no período 8-18h atinge consistentemente níveis mais altos com tendência para atingir valores máximos extremos entre as 9-11h. Esta elevada consistência de valores muito baixos seguida de valores muito altos produzem um efeito cíclico na figura 4.7A)

e, como não deixaria de ser, esta distribuição de valores é o caso mais leptocúrtico observado.

#### 4.1.8 Amoníaco

Os valores relativos do teor atmosférico em amoníaco ( $NH_3$ , variável codificada através de “h”) encontram-se representados na figura 4.8.

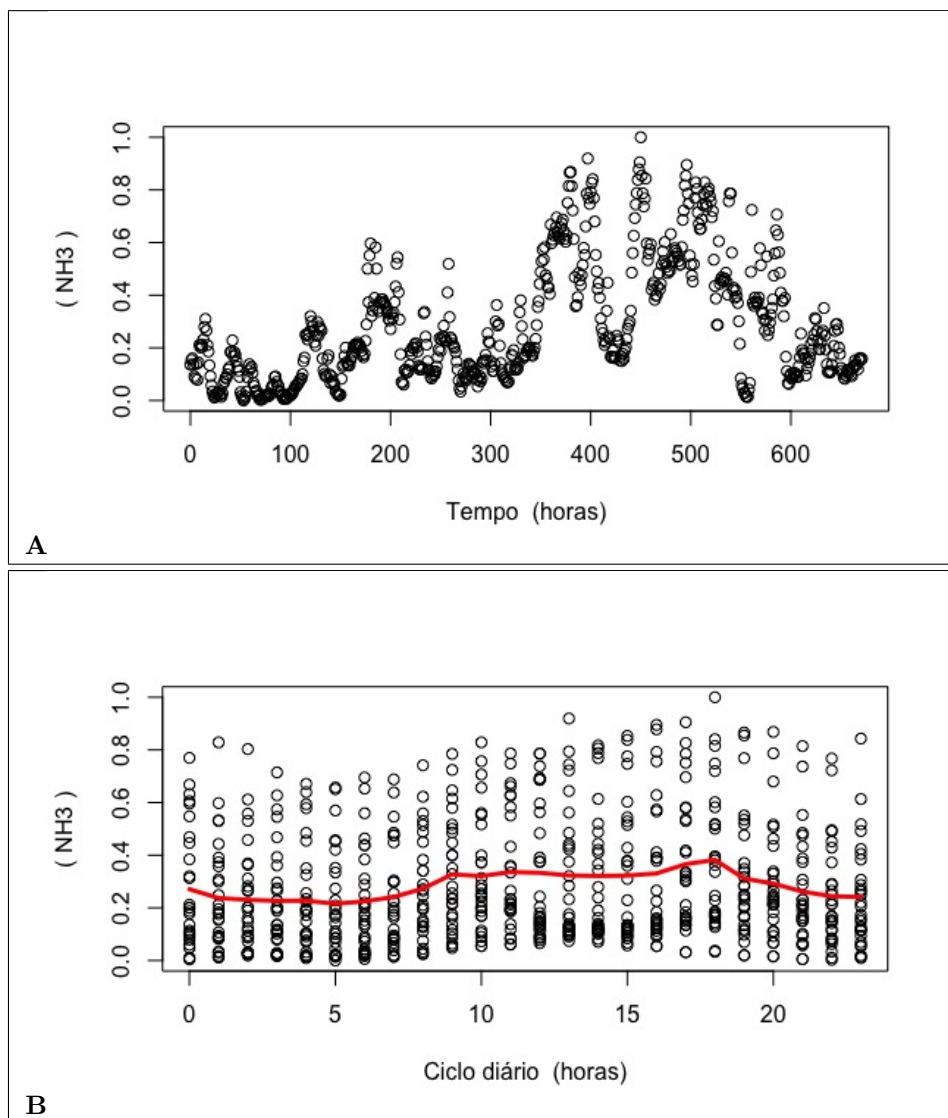


Figura 4.8: Representação dos valores relativos de amoníaco obtidos (A) durante o período de tempo considerado (em horas) e (B) segundo a periodicidade diária (escala diária de 0-23h).

Da figura 4.8 verifica-se que existe uma enorme aleatoriedade nos valores observados o que é compatível com as características mesocúrticas estimadas na



tabela 4.2.

Da figura 4.8 é difícil de evidenciar uma nítida periodicidade diária.

#### 4.1.9 NMVOCs

Os valores relativos do teor atmosférico em compostos voláteis não metânicos (NMVOCs, variável codificada através de “i”) encontram-se representados na figura 4.9.

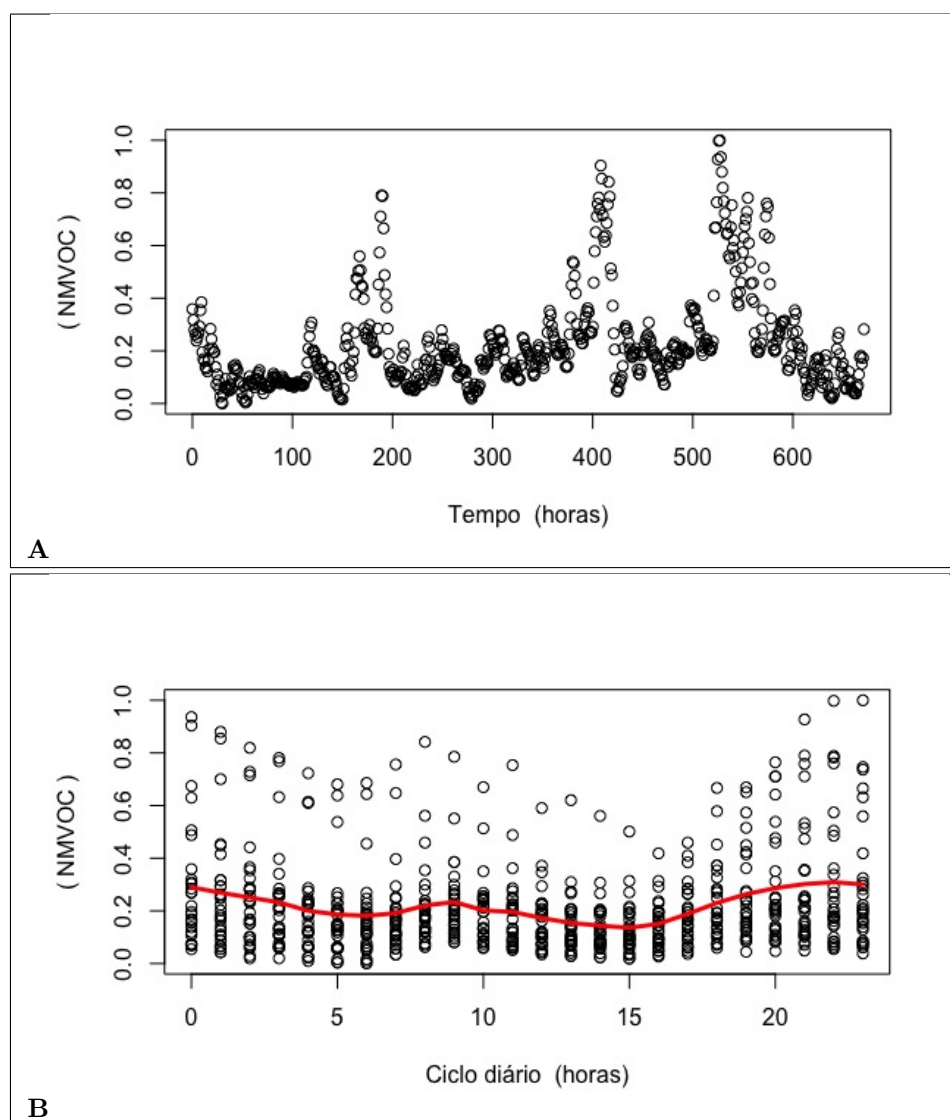


Figura 4.9: Representação dos valores relativos de NMVOCs obtidos (A) durante o período de tempo considerado (em horas) e (B) segundo a periodicidade diária (escala diária das 0-23h).

Os níveis atmosféricos de NMVOCs parecem ter uma contribuição diária

aproximadamente cíclica sobre a qual soa sobrepostas algumas situações de valores anormalmente altos. Em termos de periodicidade diária, parece haver uma sinusoidal suave com máximo entre as 20-1h e mínimo por volta das 14-16h com um máximo sobreposto entre as 7-12h.

A tendência para a consistência destes valores é compatível com as características de distribuição leptocúrtica com distorção positiva, forçada pelo conjunto de valores máximos anómalos.

#### 4.1.10 PANs

Os valores relativos do teor atmosférico em nitrato de peroxi-acetilo (PANs, variável codificada através de “j”) encontram-se representados na figura 4.10.

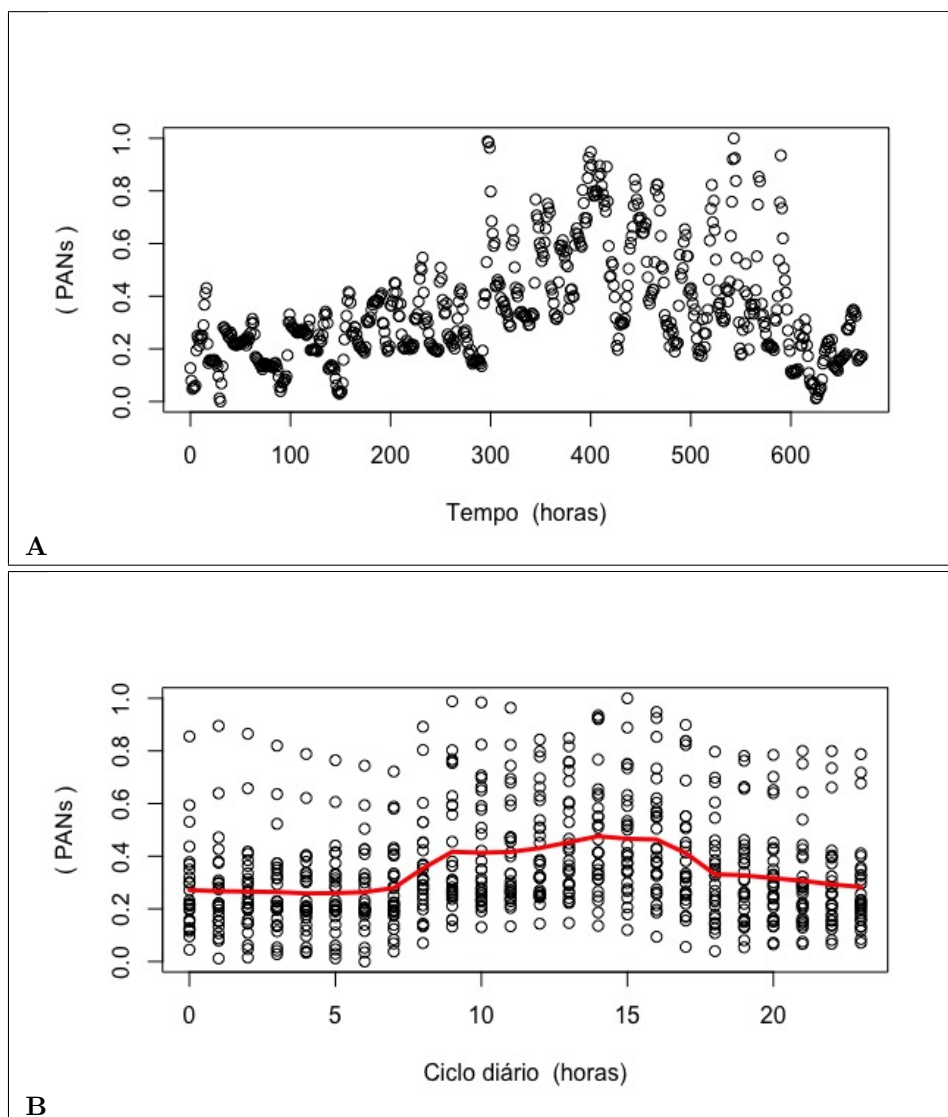


Figura 4.10: Representação dos valores relativos de PANs obtidos (A) durante o período de tempo considerado (em horas) e (B) segundo a periodicidade diária (escala diária de 0-23h).

Embora a figura 4.10A) sugira uma elevada aleatoriedade, a figura 4.10B) dá a entender que provavelmente existe uma certa periodicidade que alterna entre um nível basal mínimo no período entre as 19-7h seguido de um nível máximo entre as 9-17h.

Apesar destas evidências, devido à sua elevada aleatoriedade, esta variável apresenta-se como sendo mesocúrtica.

## 4.2 Modelação multivariada

Dado que estamos interessados em encontrar preditores para as diferentes variáveis referentes a parâmetros típicos de contaminação da qualidade do ar, decidiu-se modelar cada uma das variáveis 10 de avaliação da qualidade do ar em função das restantes 9 e respectivas interacções, incluindo as variáveis de temperatura, humidade, precipitação e velocidade do vento, num total de 79 preditores para cada resposta.

Dado que a estratégia delineada será algo repetitiva teve-se o cuidado de procurar facilitar a leitura e não ser fastidioso. Neste sentido, vamos apresentar dois casos em detalhe e abreviar os seguintes oito para que se possa chegar a conclusões válidas sem ser demasiado pesado na transmissão destes resultados.

Com foi anteriormente referido no capítulo 3, a abordagem escolhida consistia iniciar a modelação com um modelo completo e ir progressivamente removendo parâmetros com menor impacto sobre os critérios AIC e BIC tendo sido testadas as abordagens unidireccional (Backward) e bidireccional (Backward-Forward).

### 4.2.1 Monóxido de carbono

Na figura (4.11) encontra-se representado o impacto relativo de cada variável nos modelos otimizados através de AIC e BIC.

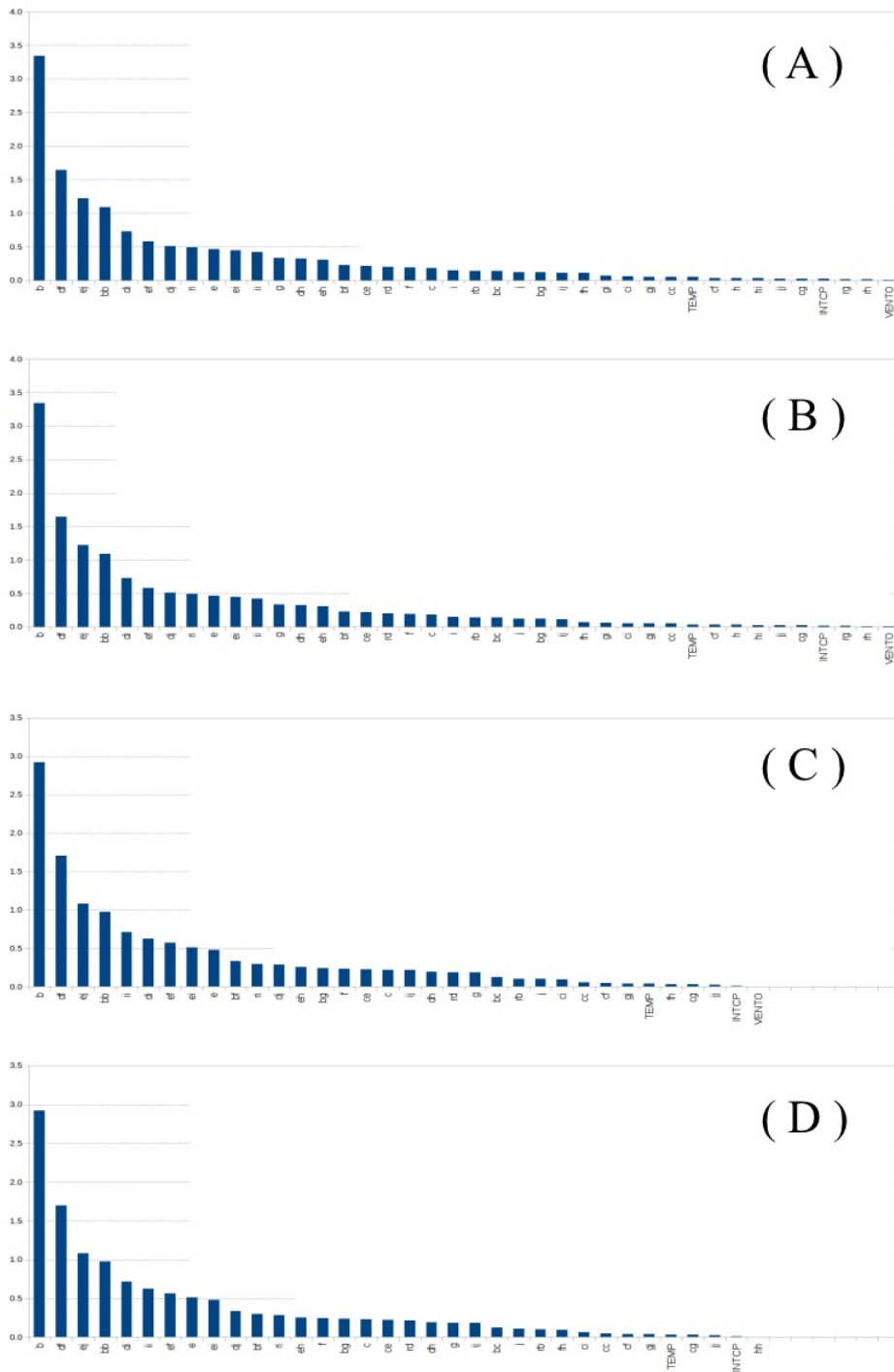


Figura 4.11: Relevância relativa (valor quadrado ordenado no sentido decrescente) dos parâmetros dos modelos otimizados obtidos com critério de Akaike (A) Backward (AIC-B), (B) Backward-Forward (AIC-BF) e critério Bayesiano (C) em modo Backward (BIC-B) e (D) em modo bidirecional (BIC-BF).

Da figura (4.11) nota-se que alguns primeiros parâmetros de cada modelo obtido têm um impacto muito maior que outros, com relevância residual.

Em cada caso de modelo refinado tem-se que decidir quais os parâmetros que são de facto relevantes.

#### 4.2.1.1 AIC-B

No sentido de averiguar qual o conjunto de parâmetros que é de facto relevante para a descrição da resposta encontrada com AIC-B, pegou-se num modelo nitidamente sobre-parametrizado (com  $p = 20$ ) e efectuou-se a remoção paulatina de um parâmetro de cada vez segundo a sequência encontrada com AIC-B, ver figura 4.11.

Na tabela 4.3 encontram-se sistematizados os valores encontrados no processo de refinação manual do modelo AIC-B.

Tabela 4.3: Evolução dos indicadores de desempenho do modelo AIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 12 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf+ce+rd+f+c | 0.00344        | -1883.3 | -1795.0 | 0.8643         |
| 19 | 13 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf+ce+rd+f   | 0.00343        | -1885.2 | -1795.0 | 0.8642         |
| 18 | 11 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf+ce+rd     | 0.00355        | -1863.5 | -1777.8 | 0.8594         |
| 17 | 8  | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf+ce        | 0.00362        | -1851.8 | -1770.6 | 0.8565         |
| 16 | 12 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf           | 0.00363        | -1851.5 | -1774.8 | 0.8560         |
| 15 | 11 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh              | 0.00364        | -1850.3 | -1778.1 | 0.8553         |
| 14 | 12 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh                 | 0.00364        | -1851.4 | -1783.8 | 0.8551         |
| 13 | 11 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g                    | 0.00375        | -1831.9 | -1768.8 | 0.8504         |
| 12 | 10 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii                      | 0.00379        | -1825.3 | -1766.7 | 0.8485         |
| 11 | 8  | b+df+ej+bb+di+ef+dj+ri+e+ei                         | 0.00386        | -1815.1 | -1761.0 | 0.8457         |
| 10 | 8  | b+df+ej+bb+di+ef+dj+ri+e                            | 0.00385        | -1816.7 | -1767.1 | 0.8456         |
| 9  | 8  | b+df+ej+bb+di+ef+dj+ri                              | 0.00388        | -1813.7 | -1768.6 | 0.8444         |
| 8  | 8  | b+df+ej+bb+di+ef+dj                                 | 0.00509        | -1631.3 | -1590.7 | 0.7953         |
| 7  | 6  | b+df+ej+bb+di+ef                                    | 0.00540        | -1593.1 | -1557.0 | 0.7827         |
| 6  | 6  | b+df+ej+bb+di                                       | 0.00539        | -1595.0 | -1563.4 | 0.7827         |
| 5  | 4  | b+df+ej+bb  | 0.00621        | -1500.7 | -1473.6 | 0.7492         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

No sentido de melhor interpretar os resultados da tabela 4.3, representou-se na figura (4.12) a evolução do desvio padrão do ajuste ( $\sigma_{fit}$ ), coeficiente de determinação (R<sup>2</sup>) e os índices de AIC e BIC em função do número de parâmetros do modelo referentes ao modelo otimizado através de AIC-B (usando o critério AIC e em modo de simplificação unidireccional).

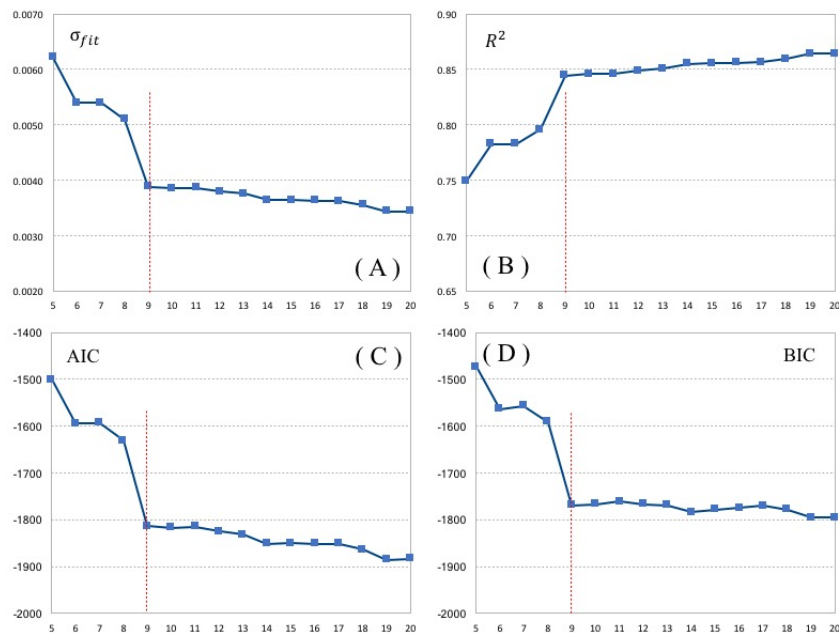


Figura 4.12: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste ( $\sigma_{fit}$ ), (B) coeficiente de determinação ( $R^2$ ), (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

Da figura (4.12) verifica-se que à medida que vai-se eliminando parâmetros a qualidade do ajuste é deteriorada (aumento do desvio padrão do ajuste e diminuição do coeficiente de determinação).

Em termos dos indicadores de desempenho AIC e BIC, as figuras 4.12(C) e 4.12(D) evidenciam a constância em BIC ao variar  $p$  entre 20 e 9 - por esta razão o critério BIC é hoje em dia considerado como mais rigoroso já que consegue melhor compensar o impacto do número de parâmetros e seu efeito aleatório no erro do modelo.

No computo global, a figura (4.12) revela uma grande alteração ao passar de  $p = 9$  para  $p = 8$ .

Na figura (4.13) apresenta-se a função de recuperação da resposta do modelo em relação ao número de parâmetros mantidos.

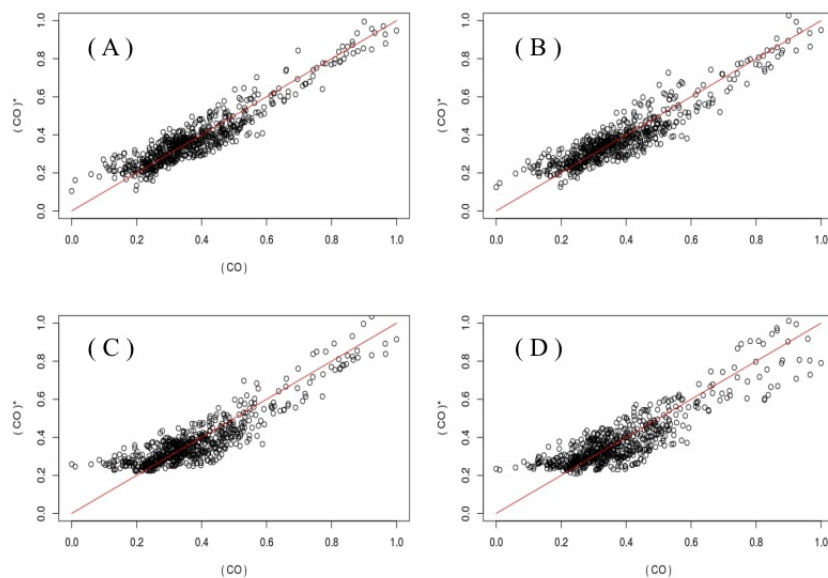


Figura 4.13: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 9$ , (C) modelo deficitário com  $p = 8$  e (D) modelo extremo com  $p = 5$ .

Da figura (4.13) verifica-se pouca diferença no modo como o modelo multivariado descreve a resposta com o número de parâmetros a variar entre  $p = 20$  (A) e  $p = 9$  (B); já a figura (C), referente a  $p = 8$ , revela uma distorção muito significativa nos valores mais baixos que é agravada também com um aumento de imprecisão ao simplificar ainda mais o modelo até obter  $p = 5$  (D).

Assim sendo o melhor modelo encontrado por AIC-B corresponde a  $p = 9$ .

Na tabela (4.4) encontram-se as estimativas paramétricas obtidas para o modelo com  $p = 9$ .

Tabela 4.4: Estimativas paramétricas do ajuste com  $p = 9$  e modelo refinado obtido por remoção do parâmetro sem significado estatístico.

| Modelo | AIC-B( $p = 9$ ) |       |       | AIC-B( $p = 8$ ) |       |       |
|--------|------------------|-------|-------|------------------|-------|-------|
|        | Estimativa       | SE    | p[H0] | Estimativa       | SE    | p[H0] |
| INTCP. | 0.111            | 0.010 | 0.000 | 0.112            | 0.010 | 0.000 |
| b      | 0.262            | 0.056 | 0.000 | 0.326            | 0.034 | 0.000 |
| df     | <b>-1.235</b>    | 0.133 | 0.000 | <b>-1.183</b>    | 0.128 | 0.000 |
| ej     | -0.316           | 0.096 | 0.001 | -0.286           | 0.094 | 0.002 |
| bb     | 0.108            | 0.076 | 0.155 | —                | —     | —     |
| di     | 0.391            | 0.060 | 0.000 | 0.428            | 0.055 | 0.000 |
| ef     | 0.540            | 0.096 | 0.000 | 0.496            | 0.091 | 0.000 |
| dj     | <b>0.874</b>     | 0.129 | 0.000 | <b>0.834</b>     | 0.126 | 0.000 |
| ri     | 0.461            | 0.032 | 0.000 | 0.440            | 0.028 | 0.000 |

SE - ); p[H0] - valor de prova



Os resultados da tabela (4.4) revelam que existe um parâmetro sem significado estatístico - o termo de interação  $bb$  possui um valor de prova superior a 5% o que pode indicar não ser estatisticamente relevante.

O modelo encontrado foi posteriormente simplificado e os parâmetros encontrados encontram-se também representados na tabela (4.4). Na figura (4.14) encontra-se representada a recuperação da resposta com o modelo  $p = 9$  [AIC-B( $p=9$ )] e  $p = 8$  [AIC-B( $p = 8$ )].

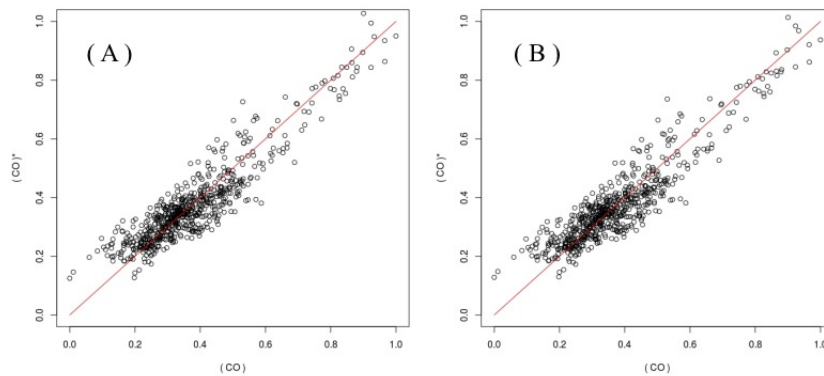


Figura 4.14: Comparação da recuperação da resposta do modelo (A) com  $p = 9$  [AIC-B( $p=9$ )] com o modelo simplificado com  $p = 8$  (B) obtido a partir do anterior, removendo o termo sem significado estatístico.

Como seria de esperar, a remoção do termo sem significado estatístico não perturbou significativamente a capacidade de recuperação da resposta: o coeficiente de determinação variou de 0.8444 para 0.8440 enquanto que o desvio padrão do ajuste passou de  $3.875 \times 10^{-3}$  para  $3.881 \times 10^{-3}$ , o que, olhando à figura 4.13 revela não ser uma variação significativa.

Além disso, atentando ainda nos valores da tabela 4.4, verifica-se que o modelo AIC-B( $p = 8$ ) possui todos os parâmetros com significado estatístico.

Assim sendo, o melhor modelo encontrado pelo método AIC-B é de facto o AIC-B( $p = 8$ ) cujas estimativas se encontram sistematizadas na tabela 4.4.

#### 4.2.1.2 AIC-BF

No sentido de averiguar qual o conjunto de parâmetros que é de facto relevante para a descrição da resposta encontrada com AIC-B, pegou-se num modelo nitidamente sobre-parametrizado (com  $p = 20$ ) e efectuou-se a remoção paulatina de um parâmetro de cada vez segundo a sequência encontrada com AIC-BF, ver figura 4.11

Tabela 4.5: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 12 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf+ce+rd+f+c | 0.00344        | -1883.3 | -1795.0 | 0.8643         |
| 19 | 13 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf+ce+rd+f   | 0.00343        | -1885.2 | -1795.0 | 0.8642         |
| 18 | 11 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf+ce+rd     | 0.00355        | -1863.5 | -1777.8 | 0.8594         |
| 17 | 8  | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf+ce        | 0.00362        | -1851.8 | -1770.6 | 0.8565         |
| 16 | 12 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh+bf           | 0.00363        | -1851.5 | -1774.8 | 0.8560         |
| 15 | 11 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh+eh              | 0.00364        | -1850.3 | -1778.1 | 0.8553         |
| 14 | 12 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g+dh                 | 0.00364        | -1851.4 | -1783.8 | 0.8551         |
| 13 | 11 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii+g                    | 0.00375        | -1831.9 | -1768.8 | 0.8504         |
| 12 | 10 | b+df+ej+bb+di+ef+dj+ri+e+ei+ii                      | 0.00379        | -1825.3 | -1766.7 | 0.8485         |
| 11 | 8  | b+df+ej+bb+di+ef+dj+ri+e+ei                         | 0.00386        | -1815.1 | -1761.0 | 0.8457         |
| 10 | 8  | b+df+ej+bb+di+ef+dj+ri+e                            | 0.00385        | -1816.7 | -1767.1 | 0.8456         |
| 9  | 8  | b+df+ej+bb+di+ef+dj+ri                              | 0.00388        | -1813.7 | -1768.6 | 0.8444         |
| 8  | 8  | b+df+ej+bb+di+ef+dj                                 | 0.00509        | -1631.3 | -1590.7 | 0.7953         |
| 7  | 6  | b+df+ej+bb+di+ef                                    | 0.00540        | -1593.1 | -1557.0 | 0.7827         |
| 6  | 6  | b+df+ej+bb+di                                       | 0.00539        | -1595.0 | -1563.4 | 0.7827         |
| 5  | 4  | b+df+ej+bb  | 0.00621        | -1500.7 | -1473.6 | 0.7492         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

No sentido de melhor interpretar os resultados da tabela 4.5, representou-se na figura (4.15) a evolução do desvio padrão do ajuste ( $\sigma_{fit}$ ), coeficiente de determinação ( $R^2$ ) e os índices de AIC e BIC em função do número de parâmetros do modelo referentes ao modelo otimizado através de AIC-BF (usando o critério AIC e em modo de simplificação bidirecional).

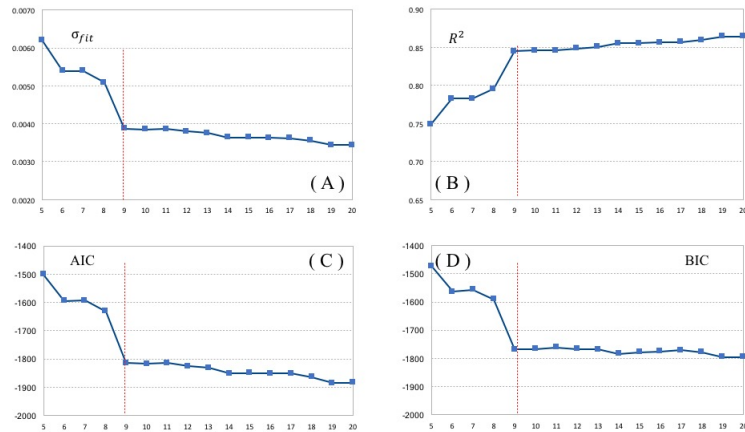


Figura 4.15: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste ( $\sigma_{fit}$ ), (B) coeficiente de determinação ( $R^2$ ), (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério de Akaike bidirecional (AIC-BF).

Da figura (4.15) verifica-se que à medida que vai-se eliminando parâmetros a qualidade do ajuste é deteriorada (aumento do desvio padrão do ajuste e

diminuição do coeficiente de determinação).

Em termos dos indicadores de desempenho AIC e BIC, as figuras 4.15(C) e 4.15(D) evidenciam a constância em BIC ao variar  $p$  entre 20 e 9 - por esta razão o critério BIC é hoje em dia considerado como mais rigoroso já que consegue melhor compensar o impacto do número de parâmetros e seu efeito aleatório no erro do modelo.

No computo global, a figura (4.15) revela uma grande alteração ao passar de  $p = 9$  para  $p = 8$ . Na figura (4.16) apresenta-se a função de recuperação da resposta do modelo em relação ao número de parâmetros mantidos.

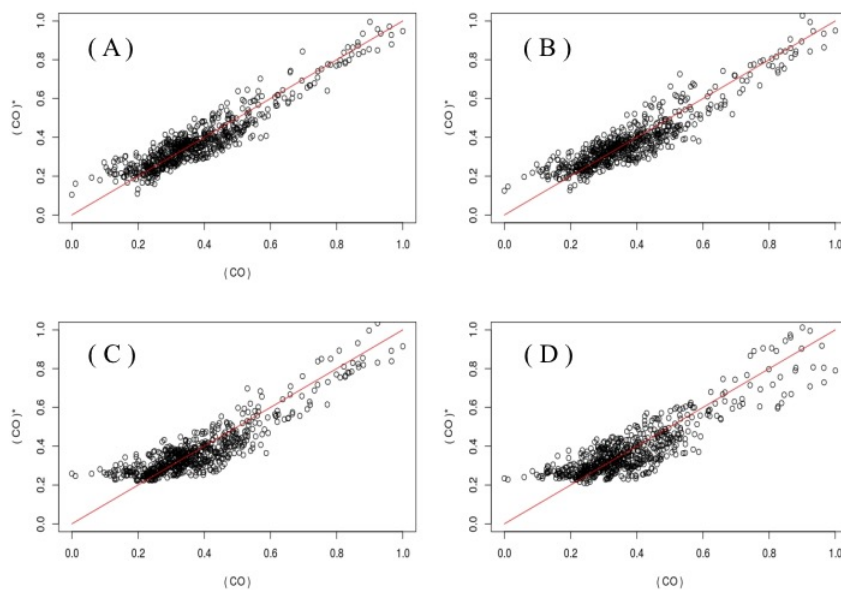


Figura 4.16: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 9$ , (C) modelo deficitário com  $p = 8$  e (D) modelo extremo com  $p = 5$ .

Da figura (4.16) verifica-se pouca diferença no modo como o modelo multivariado descreve a resposta com o número de parâmetros a variar entre  $p = 20$  (A) e  $p = 9$  (B); já a figura (C), referente a  $p = 8$ , revela uma distorção muito significativa nos valores mais baixos que é agravada também com um aumento de imprecisão ao simplificar ainda mais o modelo até obter  $p = 5$  (D). Assim sendo o melhor modelo encontrado por AIC-B corresponde a  $p = 9$ .

Na tabela (4.6) encontram-se as estimativas paramétricas obtidas para o modelo com  $p = 9$ .

Tabela 4.6: Estimativas paramétricas do ajuste com  $p = 9$  e modelo refinado obtido por remoção do parâmetro sem significado estatístico.

| Modelo | AIC-B( $p = 9$ ) |       |       | AIC-B( $p = 8$ ) |       |       |
|--------|------------------|-------|-------|------------------|-------|-------|
|        | Estimativa       | SE    | p[H0] | Estimativa       | SE    | p[H0] |
| INTCP. | 0.111            | 0.010 | 0.000 | 0.112            | 0.010 | 0.000 |
| b      | 0.262            | 0.056 | 0.000 | 0.326            | 0.034 | 0.000 |
| df     | <b>-1.235</b>    | 0.133 | 0.000 | <b>-1.183</b>    | 0.128 | 0.000 |
| ej     | -0.316           | 0.096 | 0.001 | -0.286           | 0.094 | 0.002 |
| bb     | 0.108            | 0.076 | 0.155 | —                | —     | —     |
| di     | 0.391            | 0.060 | 0.000 | 0.428            | 0.055 | 0.000 |
| ef     | 0.540            | 0.096 | 0.000 | 0.496            | 0.091 | 0.000 |
| dj     | <b>0.874</b>     | 0.129 | 0.000 | <b>0.834</b>     | 0.126 | 0.000 |
| ri     | 0.461            | 0.032 | 0.000 | 0.440            | 0.028 | 0.000 |

SE - Erro padrão (estimativa da dispersão do parâmetro); p[H0] - valor de prova

Os resultados da tabela (4.6) revelam que existe um parâmetro sem significado estatístico - o termo de interação bb possui um valor de prova superior a 5% o que pode indicar não ser estatisticamente relevante. O modelo encontrado foi posteriormente simplificado e os parâmetros encontrados encontram-se também representados na tabela (4.6). Na figura (4.17) encontra-se representada a recuperação da resposta com o modelo  $p = 9$  [AIC-B( $p=9$ )] e  $p = 8$  [AIC-B( $p = 8$ )].

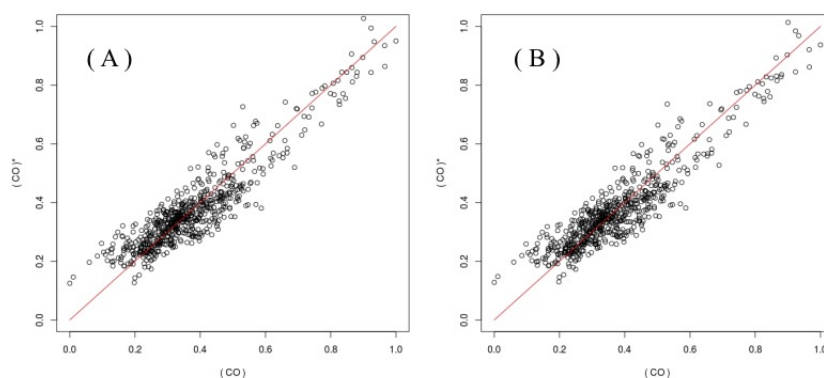


Figura 4.17: Comparação da recuperação da resposta do modelo (A) com  $p = 9$  [AIC-B( $p=9$ )] com o modelo simplificado com  $p = 8$  (B) obtido a partir do anterior, removendo o termo sem significado estatístico.

Como seria de esperar, a remoção do termo sem significado estatístico não perturbou significativamente a capacidade de recuperação da resposta: o coeficiente de determinação variou de 0.8444 para 0.8440 enquanto que o desvio padrão do ajuste passou de  $3.875 \times 10^{-3}$  para  $3.881 \times 10^{-3}$ , o que, olhando à figura 4.16 revela não ser uma variação significativa. Além disso, atentando ainda nos valores da tabela 4.6, verifica-se que o modelo AIC-B( $p = 8$ ) possui todos os parâmetros com significado estatístico. Assim sendo, o melhor modelo

encontrado pelo método AIC-B é de facto o AIC-B( $p = 8$ ) cujas estimativas se encontram sistematizadas na tabela 4.6.

#### 4.2.1.3 BIC-B

Recorrendo agora ao critério Bayesiano em modo simplificador (BIC-B), os resultados obtidos por refinamento do modelo BIC-B, na figura 4.11.

Tabela 4.7: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 6  | b+df+ej+bb+ii+di+ef+ei+e+bf+ri+dj+eh+bg+f+ce+c+ij+dh | 0.00360        | -1851.6 | -1756.9 | 0.8577         |
| 19 | 7  | b+df+ej+bb+ii+di+ef+ei+e+bf+ri+dj+eh+bg+f+ce+c+ij    | 0.00360        | -1853.2 | -1763.0 | 0.8576         |
| 18 | 7  | b+df+ej+bb+ii+di+ef+ei+e+bf+ri+dj+eh+bg+f+ce+c       | 0.00361        | -1852.0 | -1766.3 | 0.8569         |
| 17 | 8  | b+df+ej+bb+ii+di+ef+ei+e+bf+ri+dj+eh+bg+f+ce         | 0.00361        | -1854.0 | -1772.8 | 0.8569         |
| 16 | 8  | b+df+ej+bb+ii+di+ef+ei+e+bf+ri+dj+eh+bg+f            | 0.00362        | -1853.3 | -1776.7 | 0.8563         |
| 15 | 11 | b+df+ej+bb+ii+di+ef+ei+e+bf+ri+dj+eh+bg              | 0.00363        | -1851.6 | -1779.4 | 0.8555         |
| 14 | 11 | b+df+ej+bb+ii+di+ef+ei+e+bf+ri+dj+eh                 | 0.00363        | -1853.5 | -1785.9 | 0.8555         |
| 13 | 9  | b+df+ej+bb+ii+di+ef+ei+e+bf+ri+dj                    | 0.00374        | -1833.7 | -1770.5 | 0.8507         |
| 12 | 10 | b+df+ej+bb+ii+di+ef+ei+e+bf+ri                       | 0.00380        | -1824.8 | -1766.2 | 0.8483         |
| 11 | 8  | b+df+ej+bb+ii+di+ef+ei+e+bf                          | 0.00466        | -1687.2 | -1633.0 | 0.8133         |
| 10 | 7  | b+df+ej+bb+ii+di+ef+ei+e                             | 0.00491        | -1653.3 | -1603.7 | 0.8030         |
| 9  | 7  | b+df+ej+bb+ii+di+ef+ei                               | 0.00518        | -1618.7 | -1573.6 | 0.7920         |
| 8  | 7  | b+df+ej+bb+ii+di+ef                                  | 0.00523        | -1613.2 | -1572.6 | 0.7897         |
| 7  | 7  | b+df+ej+bb+ii+di                                     | 0.00524        | -1613.3 | -1577.2 | 0.7891         |
| 6  | 6  | b+df+ej+bb+ii  | 0.00531        | -1604.8 | -1573.2 | 0.7858         |
| 5  | 4  | b+df+ej+bb   | 0.00621        | -1500.7 | -1473.6 | 0.74918        |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

No sentido de melhor interpretar os resultados da tabela 4.7, representou-se na figura (4.18) a evolução do desvio padrão do ajuste ( $\sigma_{fit}$ ), coeficiente de determinação (R<sup>2</sup>) e os índices de AIC e BIC em função do número de parâmetros do modelo referentes ao modelo otimizado através de BIC-B (usando o critério BIC e em modo de simplificação unidireccional).

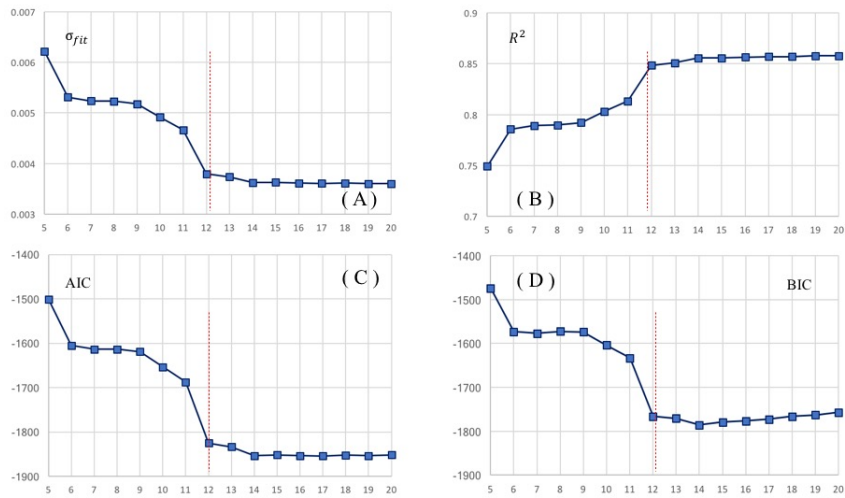


Figura 4.18: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste ( $\sigma_{fit}$ ), (B) coeficiente de determinação ( $R^2$ ), (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério Bayesiano simplificador (BIC-B).

Da figura (4.18) verifica-se que à medida que vai-se eliminando parâmetros a qualidade do ajuste é deteriorada (aumento do desvio padrão do ajuste e diminuição do coeficiente de determinação).

Em termos dos indicadores de desempenho AIC e BIC, as figuras 4.18(C) e 4.18(D) evidenciam a constância em BIC ao variar  $p$  entre 20 e 12 - por esta razão o critério BIC é hoje em dia considerado como mais rigoroso já que consegue melhor compensar o impacto do número de parâmetros e seu efeito aleatório no erro do modelo. No computo global, a figura (4.18) revela uma grande alteração ao passar de  $p = 12$  para  $p = 11$ .

Na figura (4.19) apresenta-se a função de recuperação da resposta do modelo em relação ao número de parâmetros mantidos.

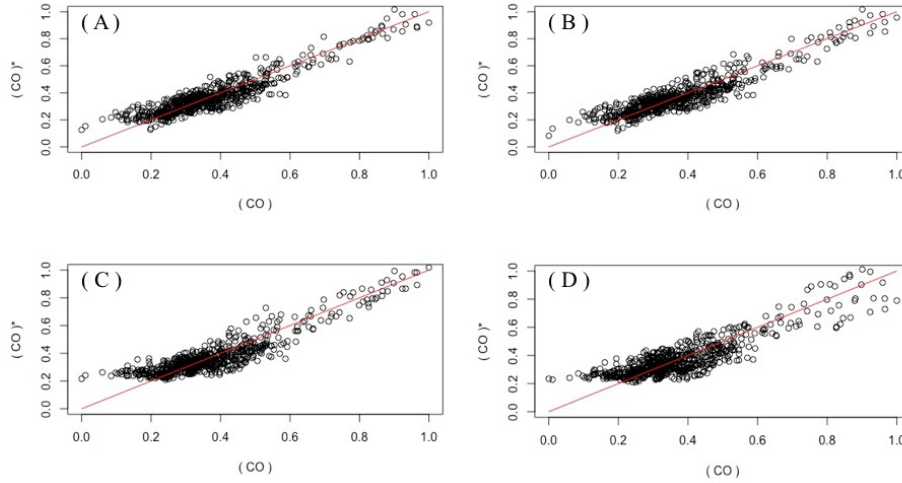


Figura 4.19: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 12$ , (C) modelo deficitário com  $p = 11$  e (D) modelo extremo com  $p = 5$ .

Da figura (4.19) verifica-se pouca diferença no modo como o modelo multivariado descreve a resposta com o número de parâmetros a variar entre  $p = 20$  (A) e  $p = 12$  (B); já a figura (C), referente a  $p = 11$ , revela uma distorção muito significativa nos valores mais baixos que é agravada também com um aumento de imprecisão ao simplificar ainda mais o modelo até obter  $p = 5$  (D). Assim sendo o melhor modelo encontrado por BIC-B corresponde a  $p = 12$ .

Tabela 4.8: Estimativas paramétricas do ajuste com  $p = 12$  e modelo refinado obtido por remoção do parâmetro sem significado estatístico.

| Modelo | BIC-B( $p = 12$ ) |         |       | BIC-B( $p = 11$ ) |       |       |
|--------|-------------------|---------|-------|-------------------|-------|-------|
|        | Estimativa        | SE      | p[H0] | Estimativa        | SE    | p[H0] |
| INTCP. | 0.103             | 0.012   | 0.000 | 0.100             | 0.007 | 0.000 |
| b      | 0.404             | 0.068   | 0.000 | 0.401             | 0.057 | 0.000 |
| df     | 0.582             | 0.201   | 0.003 | -0.711            | 0.208 | 0.000 |
| ej     | -0.297            | 0.043   | 0.000 | 0.291             | 0.047 | 0.000 |
| bb     | 0.437             | 0.115   | 0.155 | 0.441             | 0.125 | 0.000 |
| ii     | 0.248             | 0.094   | 0.008 | -0.239            | 0.084 | 0.010 |
| di     | 0.386             | 0.225   | 0.086 | 0.571             | 0.233 | 0.000 |
| ef     | 0.323             | 0.132   | 0.015 | 0.436             | 0.144 | 0.000 |
| ei     | 0.161             | 0.148   | 0.278 | -                 | -     | -     |
| e      | -0.178            | 0.03143 | 0.000 | -0.175            | 0.031 | 0.000 |
| bf     | -0.551            | 0.16401 | 0.000 | -0.561            | 0.163 | 0.000 |
| ri     | 0.514             | 0.04169 | 0.000 | 0.521             | 0.041 | 0.000 |

SE - Erro padrão (estimativa da dispersão do parâmetro); p[H0] - valor de prova

Os resultados da tabela (4.8) revelam que existe um parâmetro sem signifi-

cado estatístico - o termo de interação ei possui um valor de prova superior a 5% o que pode indicar não ser estatisticamente relevante.

O modelo encontrado foi posteriormente simplificado e os parâmetros encontrados encontram-se também representados na tabela (4.8). Na figura (4.20) encontra-se representada a recuperação da resposta com o modelo  $p = 12$  [BIC-B( $p=12$ )] e  $p = 11$  [BIC-B( $p = 11$ )].

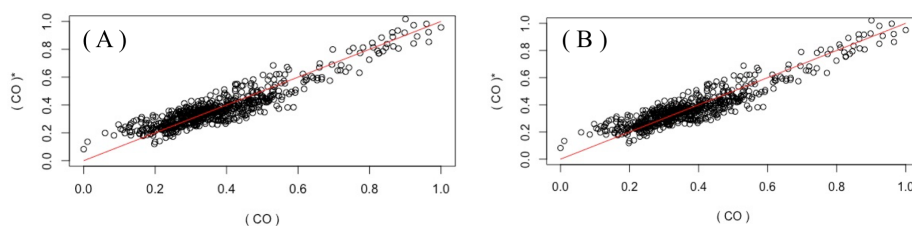


Figura 4.20: Comparação da recuperação da resposta do modelo (A) com  $p = 12$  [BIC-B( $p=9$ )] com o modelo simplificado com 11 (B) obtido a partir do anterior, removendo o termo sem significado estatístico.

Como seria de esperar, a remoção do termo sem significado estatístico não perturbou significativamente a capacidade de recuperação da resposta: o coeficiente de determinação variou de 0.8483 para 0.8480 enquanto que o desvio padrão do ajuste passou de  $3.800 \times 10^{-3}$  para  $3.796 \times 10^{-3}$ , o que, olhando à figura 4.20 revela não ser uma variação significativa. Além disso, atentando ainda nos valores da tabela 4.8, verifica-se que o modelo AIC-B( $p = 11$ ) possui todos os parâmetros com significado estatístico. Assim sendo, o melhor modelo encontrado pelo método BIC-B é de facto o AIC-B( $p = 11$ ) cujas estimativas se encontram sistematizadas na tabela 4.8.

#### 4.2.1.4 BIC-BF

No caso de se usar o critério Bayesiano em modo bidireccional (BIC-BF), pegou-se num modelo nitidamente sobre-parametrizado (com  $p = 20$ ) e efectuou-se a remoção paulatina de um parâmetro de cada vez segundo a sequência encontrada com AIC-B, ver figura 4.11.

Na tabela 4.9 encontram-se sistematizados os valores encontrados no processo de refinação manual dom modelo BIC-BF.



Tabela 4.9: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 9  | b+df+ej+bb+di+ii+ef+e+ei+dj+bf+ri+eh+f+bg+c+ce+rd+dh | 0.00345        | -1880.3 | -1785.6 | 0.8636         |
| 19 | 11 | b+df+ej+bb+di+ii+ef+e+ei+dj+bf+ri+eh+f+bg+c+ce+rd    | 0.00345        | -1882.2 | -1792.0 | 0.8636         |
| 18 | 7  | b+df+ej+bb+di+ii+ef+e+ei+dj+bf+ri+eh+f+bg+c+ce       | 0.00361        | -1852.0 | -1766.3 | 0.8569         |
| 17 | 9  | b+df+ej+bb+di+ii+ef+e+ei+dj+bf+ri+eh+f+bg+c          | 0.00361        | -1852.6 | -1771.4 | 0.8566         |
| 16 | 8  | b+df+ej+bb+di+ii+ef+e+ei+dj+bf+ri+eh+f+bg            | 0.00362        | -1853.3 | -1776.7 | 0.8563         |
| 15 | 8  | b+df+ej+bb+di+ii+ef+e+ei+dj+bf+ri+eh+f               | 0.00361        | -1855.3 | -1783.2 | 0.8563         |
| 14 | 11 | b+df+ej+bb+di+ii+ef+e+ei+dj+bf+ri+eh                 | 0.00363        | -1853.5 | -1785.9 | 0.8555         |
| 13 | 9  | b+df+ej+bb+di+ii+ef+e+ei+dj+bf+ri                    | 0.00374        | -1833.7 | -1770.5 | 0.8507         |
| 12 | 8  | b+df+ej+bb+di+ii+ef+e+ei+dj+bf                       | 0.00449        | -1711.4 | -1652.8 | 0.8204         |
| 11 | 9  | b+df+ej+bb+di+ii+ef+e+ei+dj                          | 0.00470        | -1682.7 | -1628.6 | 0.8120         |
| 10 | 7  | b+df+ej+bb+di+ii+ef+e+ei                             | 0.00491        | -1653.3 | -1603.7 | 0.8030         |
| 9  | 9  | b+df+ej+bb+di+ii+ef+e                                | 0.00498        | -1644.6 | -1599.5 | 0.7999         |
| 8  | 7  | b+df+ej+bb+di+ii+ef                                  | 0.00523        | -1613.2 | -1572.6 | 0.7897         |
| 7  | 7  | b+df+ej+bb+di+ii                                     | 0.00524        | -1613.3 | -1577.2 | 0.7891         |
| 6  | 6  | b+df+ej+bb+di  | 0.00539        | -1595.0 | -1563.4 | 0.7826         |
| 5  | 4  | b+df+ej+bb   | 0.00621        | -1500.7 | -1473.6 | 0.7491         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

No sentido de melhor interpretar os resultados da tabela 4.9, representou-se na figura (4.21) a evolução do desvio padrão do ajuste ( $\sigma_{fit}$ ), coeficiente de determinação ( $R^2$ ) e os índices de AIC e BIC em função do número de parâmetros do modelo referentes ao modelo otimizado através de BIC-BF (usando o critério BIC e em modo de simplificação bidireccional).

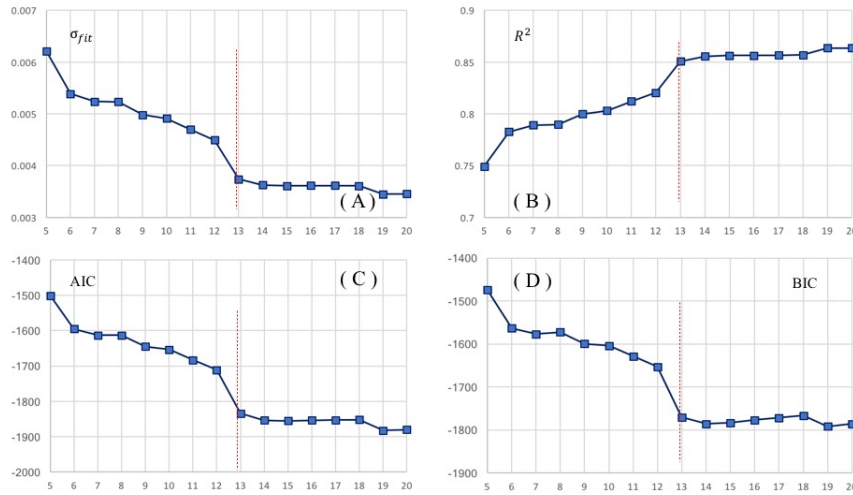


Figura 4.21: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste ( $\sigma_{fit}$ ), (B) coeficiente de determinação ( $R^2$ ), (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério bayesiano bidireccional (BIC-BF).

Da figura (4.21) verifica-se que à medida que vai-se eliminando parâmetros a qualidade do ajuste é deteriorada (aumento do desvio padrão do ajuste e diminuição do coeficiente de determinação).

Em termos dos indicadores de desempenho AIC e BIC, as figuras 4.12(C) e 4.12(D) evidenciam a constância em BIC ao variar  $p$  entre 20 e 13 - por esta razão o critério BIC é hoje em dia considerado como mais rigoroso já que consegue melhor compensar o impacto do número de parâmetros e seu efeito aleatório no erro do modelo.

No computo global, a figura (4.21) revela uma grande alteração ao passar de  $p = 13$  para  $p = 12$ . Na figura (4.22) apresenta-se a função de recuperação da resposta do modelo em relação ao número de parâmetros mantidos.

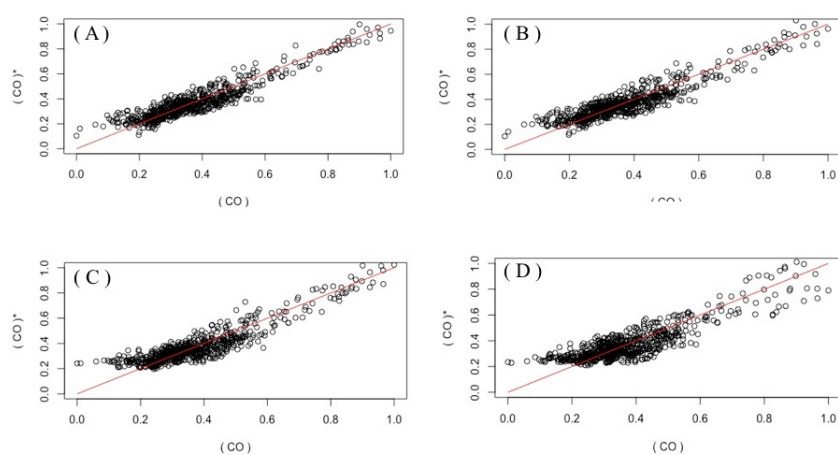


Figura 4.22: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 13$ , (C) modelo deficitário com  $p = 12$  e (D) modelo extremo com  $p = 5$ .

Da figura (4.22) verifica-se pouca diferença no modo como o modelo multivariado descreve a resposta com o número de parâmetros a variar entre  $p = 20$  (A) e  $p = 13$  (B); já a figura (C), referente a  $p = 12$ , revela uma distorção muito significativa nos valores mais baixos que é agravada também com um aumento de imprecisão ao simplificar ainda mais o modelo até obter  $p = 5$  (D). Assim sendo o melhor modelo encontrado por BIC-BF corresponde a  $p = 13$ . Na tabela (4.4) encontram-se as estimativas paramétricas obtidas para o modelo com  $p = 13$ .

Tabela 4.10: Estimativas paramétricas do ajuste com  $p = 13$  e modelo refinado obtido por remoção do parâmetro sem significado estatístico.

| Modelo | BIC-BF( $p = 13$ ) |       |       | BIC-BF( $p = 12$ ) |       |       | BIC-BF( $p = 11$ ) |       |       | BIC-BF( $p = 10$ ) |       |       |
|--------|--------------------|-------|-------|--------------------|-------|-------|--------------------|-------|-------|--------------------|-------|-------|
|        | Estimativa         | SE    | p[H0] | Estimativa         | SE    | p[H0] | Estimativa         | SE    | p[H0] | Estimativa         | SE    | p[H0] |
| INTCP. | 0.107              | 0.012 | 0.000 | 0.107              | 0.011 | 0.000 | 0.108              | 0.011 | 0.000 | 0.113              | 0.011 | 0.000 |
| b      | 0.369              | 0.069 | 0.000 | 0.373              | 0.067 | 0.000 | 0.397              | 0.06  | 0.000 | 0.438              | 0.058 | 0.000 |
| df     | -0.733             | 0.205 | 0.000 | -0.690             | 0.144 | 0.000 | -0.591             | 0.113 | 0.000 | -0.578             | 0.113 | 0.000 |
| ej     | -0.128             | 0.137 | 0.347 | -0.142             | 0.128 | 0.268 | -                  | -     | -     | -                  | -     | -     |
| bb     | 0.451              | 0.114 | 0.000 | 0.461              | 0.109 | 0.000 | 0.461              | 0.109 | 0.000 | 0.468              | 0.109 | 0.000 |
| di     | 0.071              | 0.243 | 0.768 | -                  | -     | -     | -                  | -     | -     | -                  | -     | -     |
| ii     | -0.136             | 0.099 | 0.170 | -0.120             | 0.082 | 0.144 | -0.131             | 0.081 | 0.107 | -                  | -     | -     |
| ef     | 0.406              | 0.134 | 0.002 | 0.383              | 0.109 | 0.000 | 0.320              | 0.094 | 0.000 | 0.376              | 0.087 | 0.000 |
| e      | -0.090             | 0.041 | 0.028 | -0.086             | 0.039 | 0.027 | -0.115             | 0.028 | 0.000 | -0.107             | 0.028 | 0.000 |
| ei     | 0.279              | 0.151 | 0.066 | 0.313              | 0.097 | 0.001 | 0.326              | 0.096 | 0.000 | 0.200              | 0.056 | 0.000 |
| dj     | 0.582              | 0.177 | 0.001 | 0.602              | 0.163 | 0.000 | 0.432              | 0.054 | 0.000 | 0.444              | 0.054 | 0.000 |
| bf     | -0.521             | 0.163 | 0.000 | -0.539             | 0.151 | 0.000 | -0.566             | 0.149 | 0.000 | -0.667             | 0.135 | 0.000 |
| ri     | 0.488              | 0.042 | 0.000 | 0.483              | 0.038 | 0.000 | 0.481              | 0.038 | 0.000 | 0.450              | 0.033 | 0.000 |

SE - Erro padrão (estimativa da dispersão do parâmetro); p[H0] - valor de prova

Os resultados da tabela (4.10) revelam que existem vários parâmetros sem significado estatístico - os termos de interação di, ej e o termo de segundo grau ii possui um valor de prova superior a 5% o que podem indicar não serem estatisticamente relevantes.

O modelo encontrado foi posteriormente simplificado e os parâmetros encontrados encontram-se também representados na tabela (4.10). Na figura (4.23) encontra-se representada a recuperação da resposta com o modelo  $p = 13$  [BIC-BF( $p=13$ )] e  $p = 10$  [BIC-BF( $p = 10$ )].

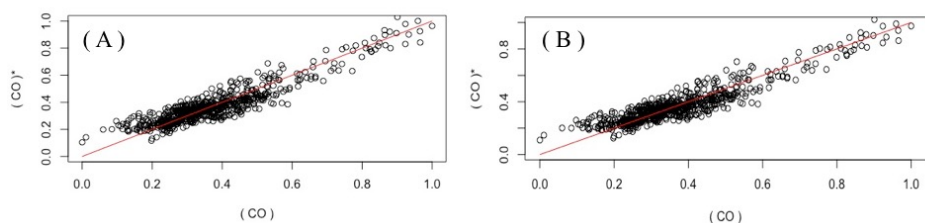


Figura 4.23: Comparação da recuperação da resposta do modelo (A) com  $p = 13$  [BIC-BF( $p=13$ )] com o modelo simplificado com  $p = 10$  (B) obtido a partir do anterior, removendo o termo sem significado estatístico.

Como seria de esperar, a remoção dos termos sem significado estatístico não perturbou significativamente a capacidade de recuperação da resposta: o coeficiente de determinação variou de 0.8507 para 0.8849 enquanto que o desvio padrão do ajuste passou de  $3.740 \times 10^{-3}$  para  $3.745 \times 10^{-3}$ , o que, olhando à figura 4.23 revela não ser uma variação significativa. Além disso, atentando ainda nos valores da tabela 4.10, verifica-se que o modelo AIC-B( $p = 10$ ) possui todos os parâmetros com significado estatístico. Assim sendo, o melhor modelo

encontrado pelo método BIC-BF é de facto o AIC-B( $p = 10$ ) cujas estimativas se encontram sistematizadas na tabela 4.10.

#### 4.2.1.5 Interpretação

De acordo com os resultados obtidos na tabela 4.11.

Tabela 4.11: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de monóxido de carbono observado durante o mês de fevereiro de 2017.

| Processo | p  | $\sigma_{fit}$ | $R^2$  | AIC     | BIC     | modelo   |
|----------|----|----------------|--------|---------|---------|--|
| AIC-B    | 8  | 0.00388        | 0.8440 | -1813.2 | -1766.9 | -1.18 df + 0.83 dj + 0.49 ef + 0.44 ri<br>+ 0.42 di + 0.32 b -0.28 ej + 0.11               |
| AIC-BF   | 8  | 0.00388        | 0.8440 | -1813.2 | -1766.9 | -1.18 df + 0.83 dj + 0.49 ef + 0.440ri<br>+ 0.42 di + 0.32 b - 0.28 ej + 0.11              |
| BIC-B    | 12 | 0.00380        | 0.8484 | -1825.6 | -1771.4 | -0.71 df +0.57 di-0.56 bf+0.52 ri+0.44 bb<br>+ 0.43 ef+0.40 b+0.29 ej-0.23 ii -0.17 e+0.10 |
| BIC-BF   | 10 | 0.00374        | 0.8498 | -1835.7 | -1786.1 | -0.66 bf-0.57 df-0.46 bb+0.45 ri+0.44 dj<br>+ 0.43 b+0.37 ef+0.20 ei-0.10 e -0.11          |

b - NO<sub>2</sub>; c - O<sub>3</sub>; d - PM<sub>2.5</sub>; e - PM<sub>10</sub>; f - SO<sub>2</sub>; g - NO;

h - NH<sub>3</sub>; i - NMVOCs; j - PANs; r? - raiz quadrada da variável ?;

Avaliando os modelos, de acordo com os processos estabelecidos para o teor de monóxido de carbono na tabela 4.11, constatamos que a otimização de modelos através dos critérios AIC-B e AIC-BF são coerentes já que permitem obter o mesmo modelo. Já a otimização de acordo com BIC-B e BIC-BF não é coerente entre elas nem com as anteriores sendo obtidos um total de três modelos.

Analisando em detalhe cada um dos casos, os resultados da tabela 4.11 evidenciam que os modelos refinado através de BIC apresentam cerca de 0.3-0.4% de maior justificação dos valores da resposta a ser modelada. Em contrapartida, estes modelos contêm mais 2 (BIC-BF) e mais 3 parâmetros que os modelos obtidos com AIC (AIC-B e AIC-BF).

Se se atender aos modelos obtidos com AIC (AIC-B e AIC-BF), verifica-se que os dois primeiros parâmetros são os mais relevantes evidenciando que as dependências do modelo otimizado com “df” e com “dj” são as mais relevantes; já os quatro seguintes parâmetros (que relacionam a resposta com as variáveis “ef”, “ri”, “di” e “b”) são relativamente importantes mas com menor impacto no modelo que as anteriores.

O termo mais relevante relaciona a resposta (níveis de CO) com “df” (variável que revela a interdependência de PM<sub>2.5</sub> com SO<sub>2</sub>) - esta interdependência está patente em todos os modelos na tabela 4.11 e assume sempre um valor negativo sugerindo que os níveis de CO aumentam quando PM<sub>2.5</sub> e SO<sub>2</sub> apresentam-se em fases opostas de variabilidade (mínimo de PM<sub>2.5</sub> com máximos de SO<sub>2</sub> e vice-versa).

O segundo termo mais relevante, embora ausente no modelo otimizado com BIC-B, expressa a interdependência entre os níveis de CO com “dj” (termo de

interacção entre  $PM_{2.5}$  e PANs) - dado que esta interdependência é positiva, espera-se obter máximos de CO quando tanto  $PM_{2.5}$  como PANs são máximos e valores mínimos quando estes são mínimos, em simultâneo.

Da combinação destes dois termos de interacção mais significativos resulta que maiores teores de CO atmosférico estão relacionados com maiores níveis de  $PM_{2.5}$  e PANs em simultâneo com um abaixamento dos níveis de  $SO_2$ .

Dos restantes 4 termos ainda relevantes, todos são positivos. Os termos “ef”, “ri” e “b” coexistem em todos os modelos otimizados enquanto que o termo “di” está ausente no modelo otimizado através de BIC-BF.

A interdependência “ef” traduz um efeito construtivo entre  $PM_{10}$  e  $SO_2$  nos níveis de CO - maiores níveis de  $PM_{10}$  e de  $SO_2$  estão relacionados com maiores teores de CO.

As variáveis “ri” (raiz quadrada do níveis de NMVOCs) e “b” ( $NO_2$ ) são também positivas demonstrando um efeito “não linear” e “linear” (directo) sobre os níveis de CO.

O termo “di”, embora ausente no modelo obtido com BIC-BF, revela a contribuição positiva de  $PM_{2.5}$  e de NMVOCs nos níveis de CO - maiores níveis de  $PM_{2.5}$  e de NMVOCs estão relacionados com teores mais elevados de CO.

Em termos de química atmosférica, atendendo ao que foi dito na sub-secção 1.3.7, esperava-se detectar alguma relação entre os níveis atmosféricos de monóxido de carbono e o teor de ozono atmosférico - nenhum dos modelos refinados foi capaz de detectar esta interdependência.

Como evidências ficou que os níveis de CO atmosférico estão relacionados com:

- efeitos construtivos entre  $PM_{2.5}$ -PANs acompanhados de um decréscimo simultâneo em  $SO_2$
- interdependência construtiva de  $PM_{10}$ - $SO_2$  e de  $PM_{2.5}$ -NMVOCs
- aumento directo de NMVOCs e  $NO_2$ .

As dependências e interdependências encontradas parecem sugerir que tanto a matéria particulada ( $PM_{10}$  e  $PM_{2.5}$ ) bem como outros constituintes mais complexos (PANs e NMVOCs) tendem a aumentar com o aumento do teor em CO.

O aumento directo entre  $NO_2$  e CO pode ser explicado como gases de exaustão comuns ao processo de combustão - se se atentar nas figuras 4.1 e 4.2, os ciclos diários de CO e  $NO_2$  parecem seguir uma tendência comum, apresentado máximos relativos nas alturas de maior tráfego rodoviário e período normal de laboração.

Tentando ir um pouco mais além, em termos químicos, o monóxido de carbono é composto um átomo de oxigénio e outro de carbono, formando uma ligação tripla entre ambos, como o oxigénio é um forte atrator de electrões, uma grande polarizabilidade, vai criar um dipolo, deslocar a carga para o seu átomo, [51]viabilizando a ligação no átomo de carbono de outra espécie com carga oposta, carga negativa, sabemos que o PM podem ser formados na atmosfera através de emissões  $SO_2$ , como poluente secundário, o que pode levar devido as propriedades de ambos, em termos cargas presentes no PM, serem adsorvido por este material e transportado das zonas onde é gerado, tais como processos de combustão, sabemos que um dos produtos destas reacções é o monóxido de carbono, bem como o PM e  $SO_2$ , principalmente de reacções que provem de combustíveis fósseis, o que explica a sua forte associação no modelo anteriormente proposto, quando estas espécies  $PM_{2.5}/SO_2$  aumentam, menor

vai ser o teor de monóxido de carbono, devido as fortes interações [52] com o monóxido de carbono, o que pode explicar em relação aos  $PM_{10}/SO_2$ , fraca interacção, que pode advir da temperatura, da pressão e da área da superfície entre esta espécie, diminuindo a sua adsorvidade, aumentando assim o numero de moléculas de monóxido de carbono, se analisarmos o  $PM_{10}$  tem um tamanho maior, minimiza a polarizibilidade da molécula, daí a relação positiva no modelo, em relação ao teor de monóxido de carbono.

Falta-nos descrever a associação com o PANs/ $PM_{2.5}$  (dj), os PANs são compostos formados por espécies com grande polarização, oxigénio e azoto, e sabemos que os PANs são poluentes secundários, que participam no transporte de outras espécies tais como  $NO_X$ , que evidenciamos a sua presença no modelo pela letra “b”, daí serem considerados bons marcadores de poluição, o que pode explicar a sua interacção com  $PM_{2.5}$  e a relação com teor de monóxido de carbono, a relação é positiva em relação ao teor, devido à maior interacção entre  $PM_{2.5}$  e o PANs, reduzindo a polarizibilidade da molécula, que pode advir da temperatura, da pressão e da área da superfície entre esta espécie, enfraquecendo a adsorção, neste caso a libertação de monóxido de carbono.

#### 4.2.2 Dióxido de azoto

Na figura 4.24 encontra-se representado o impacto relativo de cada variável nos modelos otimizados através de AIC e BIC.



Na tabela 4.12 encontram-se sistematizados os valores encontrados no processo de renação manual dom modelo AIC-B.

Tabela 4.12: Evolução da qualidade do ajuste dos valores referentes à modelação dos níveis de dióxido de azoto segundo o critério de Akaike em modo simplificativo (AIC-B) em função do sub-conjunto de parâmetros utilizados.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 13 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh+ad+cj+gg+aa+ri+eh+c+j | 0.00255        | -2085.4 | -1990.7 | 0.9123         |
| 19 | 11 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh+ad+cj+gg+aa+ri+eh+c   | 0.00256        | -2081.4 | -1991.2 | 0.9115         |
| 18 | 11 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh+ad+cj+gg+aa+ri+eh     | 0.00257        | -2082.0 | -1996.3 | 0.9113         |
| 17 | 12 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh+ad+cj+gg+aa+ri        | 0.00263        | -2066.5 | -1985.3 | 0.9090         |
| 16 | 12 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh+ad+cj+gg+aa           | 0.00263        | -2068.0 | -1991.3 | 0.9089         |
| 15 | 11 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh+ad+cj+gg              | 0.00279        | -2029.1 | -1956.9 | 0.9032         |
| 14 | 11 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh+ad+cj                 | 0.00279        | -2030.2 | -1962.6 | 0.9031         |
| 13 | 11 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh+ad                    | 0.00331        | -1915.8 | -1852.6 | 0.8848         |
| 12 | 11 | d+df+ii+ai+i+dj+dg+ae+eg+ag+dh                       | 0.00332        | -1915.5 | -1856.9 | 0.8844         |
| 11 | 10 | d+df+ii+ai+i+dj+dg+ae+eg+ag                          | 0.00343        | -1893.3 | -1839.2 | 0.8802         |
| 10 | 8  | d+df+ii+ai+i+dj+dg+ae+eg                             | 0.00368        | -1847.9 | -1798.3 | 0.8714         |
| 9  | 9  | d+df+ii+ai+i+dj+dg+ae                                | 0.00368        | -1847.7 | -1802.6 | 0.8710         |
| 8  | 6  | d+df+ii+ai+i+dj+dg                                   | 0.00378        | -1832.2 | -1791.6 | 0.8676         |
| 7  | 6  | d+df+ii+ai+i+dj                                      | 0.00465        | -1692.7 | -1656.6 | 0.8365         |
| 6  | 5  | d+df+ii+ai+i   | 0.00544        | -1588.3 | -1556.8 | 0.8085         |
| 5  | 4  | d+df+ii+ai   | 0.00602        | -1522.0 | -1494.9 | 0.7880         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

No sentido de melhor interpretar os resultados da tabela 4.12, representou-se na figura ( 4.25 na próxima página) a evolução do desvio padrão do ajuste ( $\sigma_{fit}$ ), coeficiente de determinação ( $R^2$ ) e os índices de AIC e BIC em função do número de parâmetros do modelo referentes ao modelo otimizado através de AIC-B (usando o critério AIC e em modo de simplificação unidireccional).



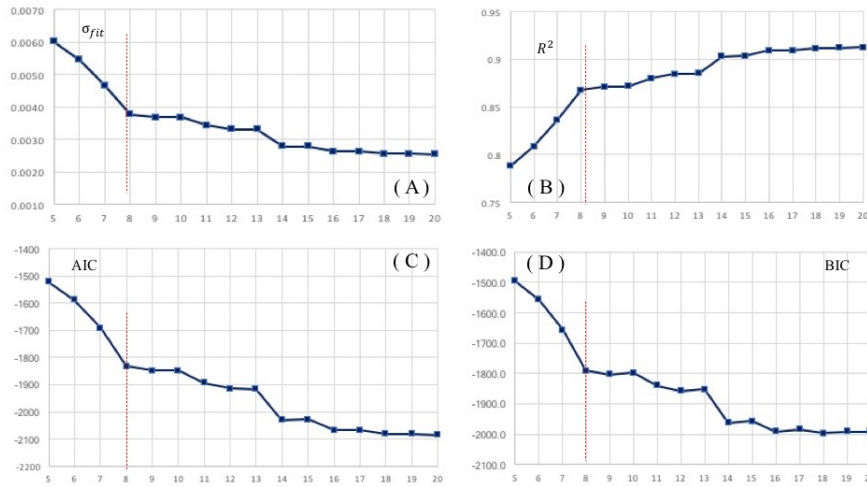


Figura 4.25: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste ( $\sigma_{fit}$ ), (B) coeficiente de determinação ( $R^2$ ), (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

Da figura ( 4.25) verifica-se que à medida que vai-se eliminando parâmetros a qualidade do ajuste é deteriorada (aumento do desvio padrão do ajuste e diminuição do coeficiente de determinação).

Em termos dos indicadores de desempenho AIC e BIC, as figuras 4.25(C) e 4.25(D) evidenciam a constância em BIC ao variar p entre 20 e 8 - por esta razão o critério BIC é hoje em dia considerado como mais rigoroso já que consegue melhor compensar o impacto do número de parâmetros e seu efeito aleatório no erro do modelo.

No computo global, a figura ( 4.25) não revela uma grande alteração ao passar de p = 8 para p = 7. Na figura ( 4.26 na página seguinte) apresenta-se a função de recuperação da resposta do modelo em relação ao número de parâmetros mantidos.

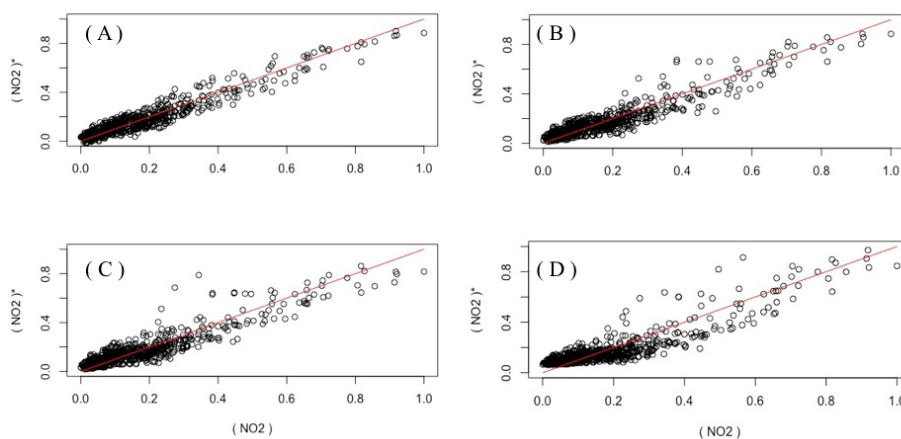


Figura 4.26: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

Da figura ( 4.26) verifica-se pouca diferença no modo como o modelo multivariado descreve a resposta com o número de parâmetros a variar entre  $p = 20$  (A) e  $p = 9$  (B); já a figura (C), referente a  $p = 8$ , revela um pequeno aumento de dispersão nos valores mais altos que é agravada também com um aumento de imprecisão ao simplificar ainda mais o modelo até obter  $p = 5$  (D). Assim sendo o melhor modelo encontrado por AIC-B corresponde a  $p = 8$ . Na tabela ( 4.13) encontram-se as estimativas paramétricas obtidas para o modelo com  $p = 8$ .

Tabela 4.13: Estimativas paramétricas do ajuste com  $p = 8$  e modelo refinado obtido por remoção do parâmetro sem significado estatístico.

| Modelo | AIC-B( $p = 8$ ) |       |       | AIC-B( $p = 7$ ) |       |       |
|--------|------------------|-------|-------|------------------|-------|-------|
|        | Estimativa       | SE    | p[H0] | Estimativa       | SE    | p[H0] |
| Intcp. | 0.011            | 0.006 | 0.074 | -                | -     | -     |
| d      | 0.114            | 0.062 | 0.067 | 0.154            | 0.058 | 0.008 |
| df     | 0.819            | 0.067 | 0.000 | 0.781            | 0.063 | 0.000 |
| ii     | -0.974           | 0.098 | 0.000 | -1.026           | 0.094 | 0.000 |
| ai     | 1.039            | 0.103 | 0.000 | 1.032            | 0.103 | 0.000 |
| i      | 0.395            | 0.047 | 0.000 | 0.449            | 0.036 | 0.000 |
| dj     | -0.523           | 0.048 | 0.000 | -0.523           | 0.048 | 0.000 |
| dg     | -0.459           | 0.036 | 0.000 | 0.466            | 0.036 | 0.000 |

SE - Erro padrão (estimativa da dispersão do parâmetro); p[H0] - valor de prova

Os resultados da tabela (4.13) revelam que existe um parâmetro sem significado estatístico - o termo de intercepção Intcp. possui um valor de prova superior a 5% o que pode indicar não ser estatisticamente relevante. O modelo encontrado foi posteriormente simplificado e os parâmetros encontrados

encontram-se também representados na tabela (4.13). Na figura (4.27) encontra-se representada a recuperação da resposta com o modelo  $p = 8$  [AIC-B( $p=8$ )] e  $p = 7$  [AIC-B( $p = 7$ )].

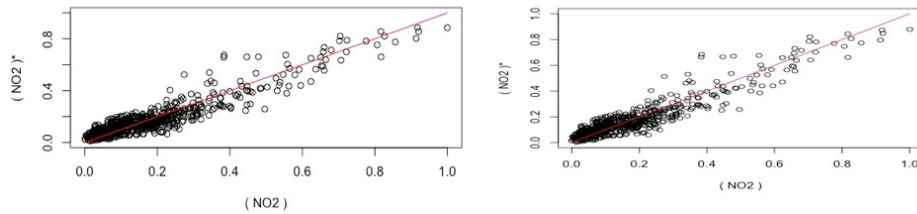


Figura 4.27: Comparação da recuperação da resposta do modelo (A) com  $p = 8$  [AIC-B( $p=8$ )] com o modelo simplificado com  $p = 7$  (B) obtido a partir do anterior, removendo o termo sem significado estatístico.

Como seria de esperar, a remoção do termo sem significado estatístico não perturbou significativamente a capacidade de recuperação da resposta: o coeficiente de determinação variou de 0.8676 para 0.8671 enquanto que o desvio padrão do ajuste passou de  $3.788 \times 10^{-3}$  para  $3.788 \times 10^{-3}$ , o que, olhando à figura 4.27 revela não ser uma variação significativa. Além disso, atentando ainda nos valores da tabela 4.13, verifica-se que o modelo AIC-B( $p = 7$ ) possui todos os parâmetros com significado estatístico. Assim sendo, o melhor modelo encontrado pelo método AIC-B é de facto o AIC-B( $p = 7$ ) cujas estimativas se encontram sistematizadas na tabela 4.13.

#### 4.2.2.2 AIC-BF

No sentido de averiguar qual o conjunto de parâmetros que é de facto relevante para a descrição da resposta encontrada com AIC-B, pegou-se num modelo nitidamente sobre-parametrizado (com  $p = 20$ ) e efectuou-se a remoção paulatina de um parâmetro de cada vez segundo a sequência encontrada com AIC-BF, ver figura 4.24.

Tabela 4.14: Evolução dos indicadores de desempenho da evolução dos valores relativos ao dióxido de azoto obtidos através do critério de Akaike bidirecional (AIC-BF) em função do sub-conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 12 | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg+ag+fg+dh+ad+ej+cj+gg+aa | 0.00217        | -2191.3 | -2096.6 | 0.9251         |
| 19 | 14 | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg+ag+fg+dh+ad+ej+cj+gg    | 0.00220        | -2184.2 | -2094.0 | 0.9241         |
| 18 | 14 | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg+ag+fg+dh+ad+ej+cj       | 0.00220        | -2185.7 | -2100.0 | 0.9240         |
| 17 | 11 | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg+ag+fg+dh+ad+ej          | 0.00254        | -2088.7 | -2007.6 | 0.9120         |
| 16 | 13 | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg+ag+fg+dh+ad             | 0.00263        | -2068.2 | -1991.6 | 0.9090         |
| 15 | 11 | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg+ag+fg+dh                | 0.00265        | -2062.9 | -1990.8 | 0.9080         |
| 14 | 9  | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg+ag+fg                   | 0.00271        | -2050.4 | -1982.7 | 0.9060         |
| 13 | 9  | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg+ag                      | 0.00297        | -1988.6 | -1925.4 | 0.8966         |
| 12 | 7  | d+ ei+df+ii+ai+i+fi+dj+dg+ae+eg                         | 0.00328        | -1923.3 | -1864.7 | 0.8857         |
| 11 | 8  | d+ ei+df+ii+ai+i+fi+dj+dg+ae                            | 0.00327        | -1924.9 | -1870.8 | 0.8857         |
| 10 | 8  | d+ ei+df+ii+ai+i+fi+dj+dg                               | 0.00348        | -1885.3 | -1835.7 | 0.8783         |
| 9  | 8  | d+ ei+df+ii+ai+i+fi+dj                                  | 0.00403        | -1788.0 | -1742.9 | 0.8590         |
| 8  | 5  | d+ ei+df+ii+ai+i+fi                                     | 0.00504        | -1638.4 | -1597.9 | 0.8233         |
| 7  | 6  | d+ ei+df+ii+ai+i  | 0.00505        | -1638.1 | -1602.0 | 0.8227         |
| 6  | 5  | d+ ei+df+ii+ai  | 0.00528        | -1609.0 | -1577.5 | 0.8143         |
| 5  | 5  | d+ ei+df+ii   | 0.00603        | -1521.0 | -1494.0 | 0.7877         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

No sentido de melhor interpretar os resultados da tabela 4.14, representou-se na figura (4.28) a evolução do desvio padrão do ajuste ( $\sigma_{fit}$ ), coeficiente de determinação ( $R^2$ ) e os índices de AIC e BIC em função do número de parâmetros do modelo referentes ao modelo otimizado através de AIC-BF (usando o critério AIC e em modo de simplificação bidirecional).

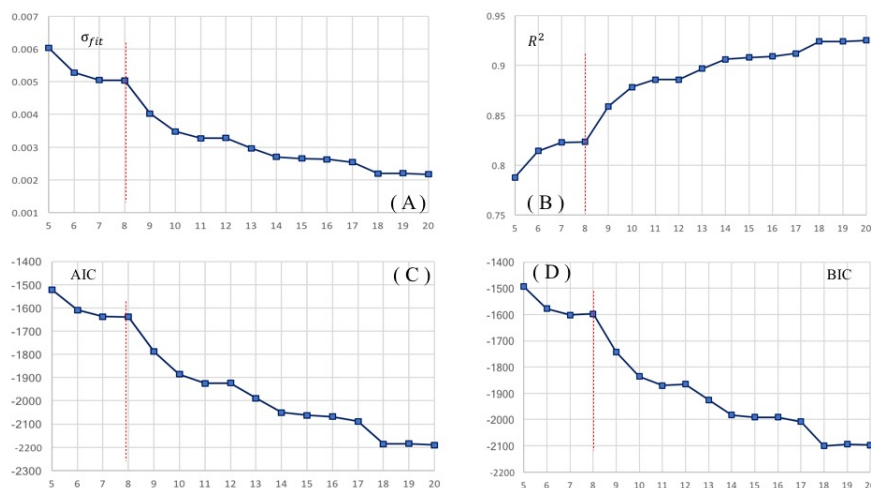


Figura 4.28: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste ( $\sigma_{fit}$ ), (B) coeficiente de determinação ( $R^2$ ), (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

Da figura (4.28) verifica-se que à medida que vai-se eliminando parâmetros a qualidade do ajuste é deteriorada (aumento do desvio padrão do ajuste e diminuição do coeficiente de determinação).

Em termos dos indicadores de desempenho AIC e BIC, as figuras 4.28(C) e 4.28(D) evidenciam um comportamento muito idêntico ao variar  $p$  entre 20 e 8 - o critério BIC neste situação não consegue compensar ou é pouco visível o impacto do número de parâmetros e seu efeito aleatório no erro do modelo.

No computo global, a figura (4.28) revela uma ligeira alteração ao passar de  $p = 8$  para  $p = 7$ .

Na figura (4.29) apresenta-se a função de recuperação da resposta do modelo em relação ao número de parâmetros mantidos

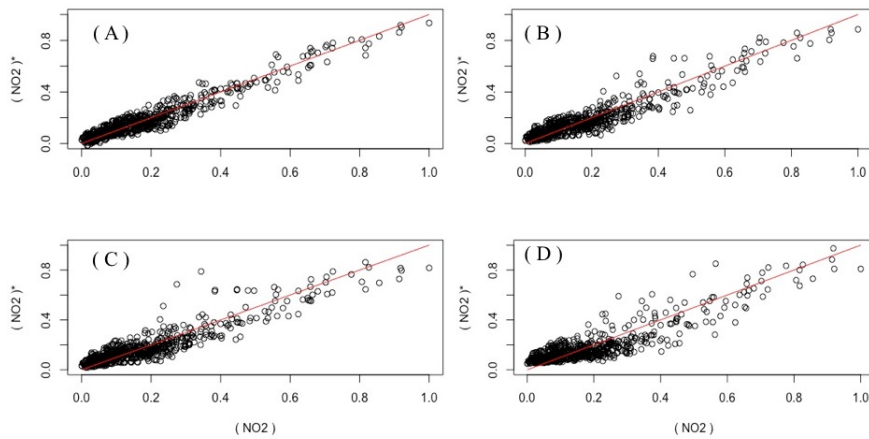


Figura 4.29: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

Da figura (4.29) verifica-se pouca diferença no modo como o modelo multivariado descreve a resposta com o número de parâmetros a variar entre  $p = 20$  (A) e  $p = 8$  (B); já a figura (C), referente a  $p = 7$ , revela uma pequena distorção não muito significativa nos valores mais baixos que é agravada também com um aumento de imprecisão ao simplificar ainda mais o modelo até obter  $p = 5$  (D). Assim sendo o melhor modelo encontrado por AIC-BF corresponde a  $p = 8$ . Na tabela (4.15) encontram-se as estimativas paramétricas obtidas para o modelo com  $p = 8$ .

Tabela 4.15: Estimativas paramétricas do ajuste com  $p = 8$  e modelo refinado obtido por remoção do parâmetro sem significado estatístico.

| Modelo | AIC-BF( $p = 8$ ) |       |       | AIC-BF( $p = 7$ ) |       |       | AIC-BF( $p = 6$ ) |       |       |
|--------|-------------------|-------|-------|-------------------|-------|-------|-------------------|-------|-------|
|        | Estimativa        | SE    | p[H0] | Estimativa        | SE    | p[H0] | Estimativa        | SE    | p[H0] |
| Intcp. | 0.017             | 0.007 | 0.019 | 0.016             | 0.007 | 0.023 | 0.019             | 0.006 | 0.004 |
| d      | 0.179             | 0.094 | 0.057 | 0.094             | 0.076 | 0.217 | -                 | -     | -     |
| ei     | -0.493            | 0.070 | 0.000 | -0.510            | 0.069 | 0.000 | -0.476            | 0.064 | 0.000 |
| df     | 0.243             | 0.137 | 0.078 | 0.426             | 0.068 | 0.000 | 0.490             | 0.045 | 0.000 |
| ii     | -0.473            | 0.139 | 0.000 | -0.358            | 0.118 | 0.002 | -0.429            | 0.103 | 0.000 |
| ai     | 1.086             | 0.129 | 0.000 | 1.009             | 0.119 | 0.000 | 1.043             | 0.115 | 0.000 |
| i      | 0.224             | 0.086 | 0.010 | 0.324             | 0.057 | 0.000 | 0.349             | 0.054 | 0.000 |
| fi     | 0.193             | 0.125 | 0.124 | -                 | -     | -     | -                 | -     | -     |

SE - Erro padrão (estimativa da dispersão do parâmetro); p[H0] - valor de prova

Os resultados da tabela (4.15) revelam que existe um parâmetro sem significado estatístico - o termo de interação fi possui um valor de prova superior a 5% o que pode indicar não ser estatisticamente relevante.

O modelo encontrado foi posteriormente simplificado e os parâmetros encontrados encontram-se também representados na tabela (4.15).

Na figura (4.30) encontra-se representada a recuperação da resposta com o modelo  $p = 8$  [AIC- BF( $p=8$ )] e  $p = 6$  [AIC-BF( $p = 6$ )].

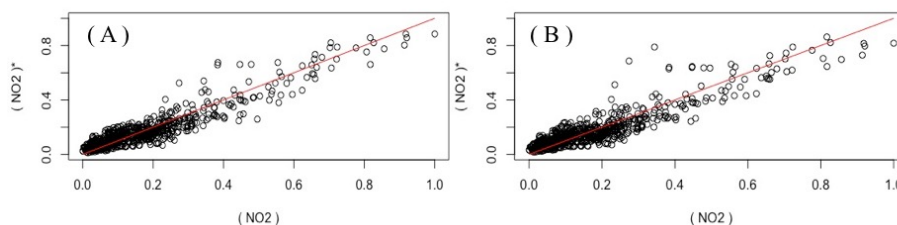


Figura 4.30: Comparação da recuperação da resposta do modelo (A) com  $p = 8$  [AIC-B( $p=8$ )] com o modelo simplificado com  $p = 6$  (B) obtido a partir do anterior, removendo os termos sem significado estatístico.

Como seria de esperar, a remoção do termo sem significado estatístico não perturbou significativamente a capacidade de recuperação da resposta: o coeficiente de determinação variou de 0.8233 para 0.8223 enquanto que o desvio padrão do ajuste passou de  $5.040 \times 10^{-3}$  para  $5.050 \times 10^{-3}$ , o que, olhando à figura 4.30 revela não ser uma variação significativa. Além disso, atentando ainda nos valores da tabela 4.15, verifica-se que o modelo AIC-BF( $p = 6$ ) possui todos os parâmetros com significado estatístico.

Assim sendo, o melhor modelo encontrado pelo método AIC-BF é de facto o AIC-BF( $p = 6$ ) cujas estimativas se encontram sistematizadas na tabela 4.15.

## 4.2.2.3 BIC-B

No sentido de averiguar qual o conjunto de parâmetros que é de facto relevante para a descrição da resposta encontrada com BIC-B, pegou-se num modelo nitidamente sobre-parametrizado (com  $p = 20$ ) e efectuou-se a remoção paulatina de um parâmetro de cada vez segundo a sequência encontrada com BIC-B, ver figura 4.24.

Na tabela 4.16 encontram-se sistematizados os valores encontrados no processo de refinação manual do modelo BIC-B.

Tabela 4.16: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 14 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh+cc+rg+cg+hh | 0.00234        | -2141.4 | -2046.7 | 0.9193         |
| 19 | 14 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh+cc+rg+cg    | 0.00246        | -2110.3 | -2020.1 | 0.9153         |
| 18 | 12 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh+cc+rg       | 0.00249        | -2102.8 | -2017.1 | 0.9140         |
| 17 | 12 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh+cc          | 0.00248        | -2104.6 | -2023.5 | 0.9140         |
| 16 | 12 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh             | 0.00248        | -2106.6 | -2030.0 | 0.9140         |
| 15 | 12 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c                | 0.00249        | -2103.7 | -2031.6 | 0.9134         |
| 14 | 9  | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi                  | 0.00324        | -1929.4 | -1861.7 | 0.8874         |
| 13 | 9  | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag                     | 0.00323        | -1931.3 | -1868.1 | 0.8874         |
| 12 | 10 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh                        | 0.00331        | -1916.1 | -1857.4 | 0.8845         |
| 11 | 10 | ei+ae+eg+dg+df+fg+i+d+ii+gg                           | 0.00333        | -1913.5 | -1859.3 | 0.8837         |
| 10 | 8  | ei+ae+eg+dg+df+fg+i+d+ii                              | 0.00341        | -1898.6 | -1848.9 | 0.8807         |
| 9  | 7  | ei+ae+eg+dg+df+fg+i+d                                 | 0.00347        | -1888.2 | -1843.1 | 0.8785         |
| 8  | 8  | ei+ae+eg+dg+df+fg+i                                   | 0.00353        | -1876.9 | -1836.3 | 0.8761         |
| 7  | 4  | ei+ae+eg+dg+df+fg                                     | 0.00613        | -1507.4 | -1471.3 | 0.7846         |
| 6  | 4  | ei+ae+eg+dg+df  | 0.00646        | -1473.1 | -1441.5 | 0.7727         |
| 5  | 3  | ei+ae+eg+dg   | 0.00884        | -1263.2 | -1236.1 | 0.6885         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

No sentido de melhor interpretar os resultados da tabela 4.16, representou-se na figura (4.31) a evolução do desvio padrão do ajuste ( $\sigma_{fit}$ ), coeficiente de determinação (R<sup>2</sup>) e os índices de AIC e BIC em função do número de parâmetros do modelo referentes ao modelo otimizado através de BIC-B (usando o critério BIC e em modo de simplificação unidireccional).

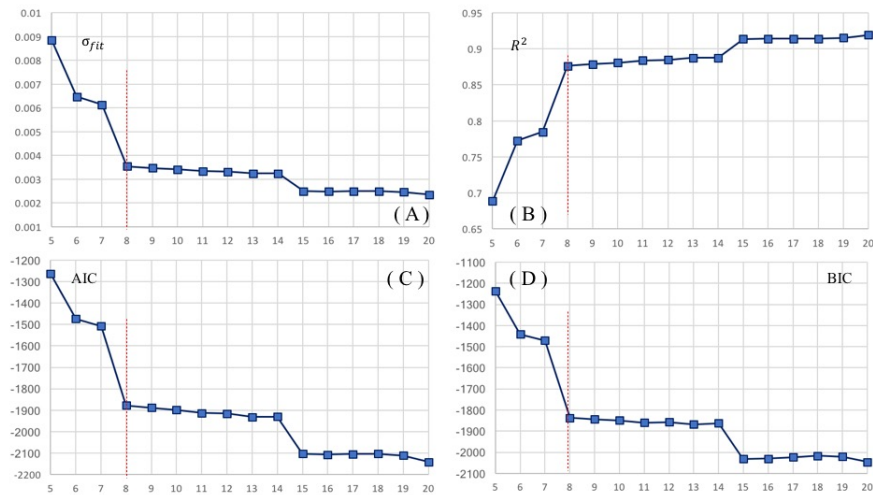


Figura 4.31: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste ( $\sigma_{fit}$ ), (B) coeficiente de determinação ( $R^2$ ), (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward.

Da figura (4.31) verifica-se que à medida que vai-se eliminando parâmetros a qualidade do ajuste é deteriorada (aumento do desvio padrão do ajuste e diminuição do coeficiente de determinação).

Em termos dos indicadores de desempenho AIC e BIC, as figuras 4.31(C) e 4.31(D) evidenciam um comportamento muito idêntico ao variar p entre 20 e 8 - o critério BIC neste situação não consegue compensar ou é pouco visível o impacto do número de parâmetros e seu efeito aleatório no erro do modelo.

No computo global, a figura (4.31) revela uma grande alteração ao passar de  $p = 8$  para  $p = 7$ .

Na figura (4.32) apresenta-se a função de recuperação da resposta do modelo em relação ao número de parâmetros mantidos



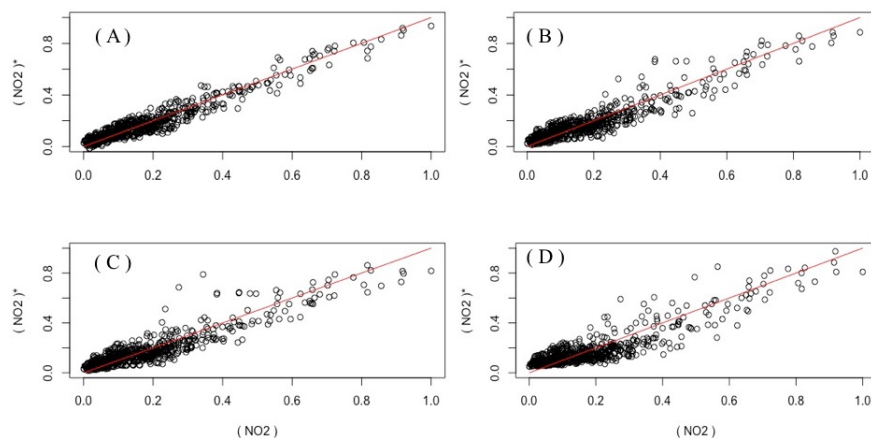


Figura 4.32: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

Da figura (4.32) verifica-se pouca diferença no modo como o modelo multivariado descreve a resposta com o número de parâmetros a variar entre  $p = 20$  (A) e  $p = 8$  (B); já a figura (C), referente a  $p = 7$ , revela uma pequena distorção não muito significativa nos valores mais baixos que é agravada também com um aumento de imprecisão ao simplificar ainda mais o modelo até obter  $p = 5$  (D). Assim sendo o melhor modelo encontrado por BIC-B corresponde a  $p = 8$ . Na tabela (4.17) encontram-se as estimativas paramétricas obtidas para o modelo com  $p = 8$ .

Tabela 4.17: Estimativas paramétricas do ajuste com  $p = 8$  e modelo refinado obtido por remoção do parâmetro sem significado estatístico.

| Modelo    | BIC-B( $p = 8$ ) |       |       |
|-----------|------------------|-------|-------|
| Parâmetro | Estimativa       | SE    | p[H0] |
| Intcp.    | -0.017           | 0.006 | 0.004 |
| ei        | -0.959           | 0.083 | 0.000 |
| ae        | 0.537            | 0.064 | 0.000 |
| eg        | 0.843            | 0.163 | 0.000 |
| dg        | -0.310           | 0.133 | 0.020 |
| df        | 0.707            | 0.040 | 0.000 |
| fg        | -0.794           | 0.066 | 0.000 |
| i         | 0.762            | 0.034 | 0.000 |

SE - Erro padrão (estimativa da dispersão do parâmetro); p[H0] - valor de prova

Os resultados da tabela (4.17) não evidenciam a existência de parâmetros sem significado estatístico.

O modelo encontrado e os parâmetros encontrados encontram-se também representados na tabela (4.17).

Este modelo apresenta o coeficiente de determinação de 0.8761 enquanto que o desvio padrão do ajuste de  $3.530 \times 10^{-3}$ . Além disso, atentando ainda nos valores da tabela 4.17, verifica-se que o modelo BIC-B( $p = 8$ ) possui todos os parâmetros com significado estatístico.

Assim sendo, o melhor modelo encontrado pelo método BIC-B é de facto o BIC-B( $p = 8$ ) cujas estimativas se encontram sistematizadas na tabela 4.17.

#### 4.2.2.4 BIC-BF

No sentido de averiguar qual o conjunto de parâmetros que é de facto relevante para a descrição da resposta encontrada com BIC-BF, pegou-se num modelo nitidamente sobre-parametrizado (com  $p = 20$ ) e efectuou-se a remoção paulatina de um parâmetro de cada vez segundo a sequência encontrada com BIC-BF, ver figura 4.24.

Na tabela 4.18 encontram-se sistematizados os valores encontrados no processo de refinação manual do modelo BIC-BF.

Tabela 4.18: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 14 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh+cc+rg+cg+hh | 0.00234        | -2141.4 | -2046.7 | 0.9193         |
| 19 | 14 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh+cc+rg+cg    | 0.00246        | -2110.3 | -2020.1 | 0.9153         |
| 18 | 12 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh+cc+rg       | 0.00249        | -2102.8 | -2017.1 | 0.9140         |
| 17 | 12 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh+cc          | 0.00248        | -2104.6 | -2023.5 | 0.9140         |
| 16 | 12 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c+eh             | 0.00248        | -2106.6 | -2030.0 | 0.9140         |
| 15 | 12 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi+c                | 0.00249        | -2103.7 | -2031.6 | 0.9134         |
| 14 | 9  | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag+gi                  | 0.00324        | -1929.4 | -1861.7 | 0.8874         |
| 13 | 9  | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh+ag                     | 0.00323        | -1931.3 | -1868.1 | 0.8874         |
| 12 | 10 | ei+ae+eg+dg+df+fg+i+d+ii+gg+dh                        | 0.00331        | -1916.1 | -1857.4 | 0.8845         |
| 11 | 10 | ei+ae+eg+dg+df+fg+i+d+ii+gg                           | 0.00333        | -1913.5 | -1859.3 | 0.8837         |
| 10 | 8  | ei+ae+eg+dg+df+fg+i+d+ii                              | 0.00341        | -1898.6 | -1848.9 | 0.8807         |
| 9  | 7  | ei+ae+eg+dg+df+fg+i+d                                 | 0.00347        | -1888.2 | -1843.1 | 0.8785         |
| 8  | 8  | ei+ae+eg+dg+df+fg+i                                   | 0.00353        | -1876.9 | -1836.3 | 0.8761         |
| 7  | 4  | ei+ae+eg+dg+df+fg                                     | 0.00613        | -1507.4 | -1471.3 | 0.7846         |
| 6  | 4  | ei+ae+eg+dg+df  | 0.00646        | -1473.1 | -1441.5 | 0.7727         |
| 5  | 3  | ei+ae+eg+dg   | 0.00884        | -1263.2 | -1236.1 | 0.6885         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

No sentido de melhor interpretar os resultados da tabela 4.18, representou-se na figura (4.33) a evolução do desvio padrão do ajuste ( $\sigma_{fit}$ ), coeficiente de determinação (R<sup>2</sup>) e os índices de AIC e BIC em função do número de parâmetros do modelo referentes ao modelo otimizado através de BIC-BF (usando o critério BIC e em modo de simplificação bidireccional).

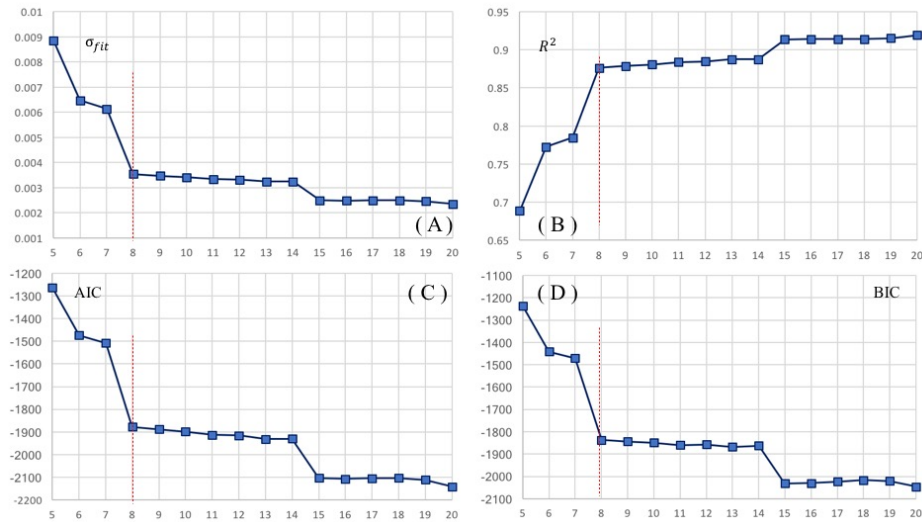


Figura 4.33: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste ( $\sigma_{fit}$ ), (B) coeficiente de determinação ( $R^2$ ), (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward-Forward.

Da figura (4.33) verifica-se que à medida que vai-se eliminando parâmetros a qualidade do ajuste é deteriorada (aumento do desvio padrão do ajuste e diminuição do coeficiente de determinação).

Em termos dos indicadores de desempenho AIC e BIC, as figuras 4.33(C) e 4.33(D) evidenciam um comportamento muito idêntico ao variar p entre 20 e 8 - o critério BIC neste situação não consegue compensar ou é pouco visível o impacto do número de parâmetros e seu efeito aleatório no erro do modelo.

No computo global, a figura (4.33) revela uma grande alteração ao passar de  $p = 8$  para  $p = 7$ .

Na figura (4.34) apresenta-se a função de recuperação da resposta do modelo em relação ao número de parâmetros mantidos

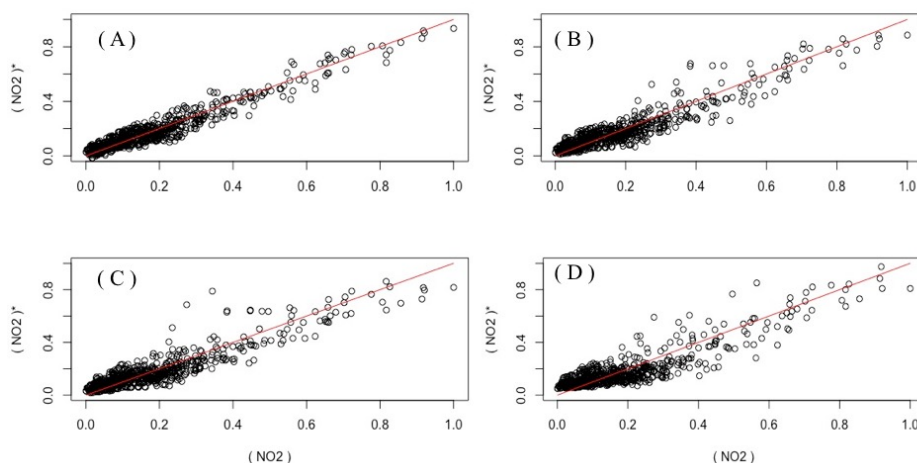


Figura 4.34: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

Da figura (4.34) verifica-se pouca diferença no modo como o modelo multivariado descreve a resposta com o número de parâmetros a variar entre  $p = 20$  (A) e  $p = 8$  (B); já a figura (C), referente a  $p = 7$ , revela uma pequena distorção não muito significativa nos valores mais baixos que é agravada também com um aumento de imprecisão ao simplificar ainda mais o modelo até obter  $p = 5$  (D). Assim sendo o melhor modelo encontrado por BIC-BF corresponde a  $p = 8$ . Na tabela (4.19) encontram-se as estimativas paramétricas obtidas para o modelo com  $p = 8$ .

Tabela 4.19: Estimativas paramétricas do ajuste com  $p = 8$  e modelo refinado obtido por remoção do parâmetro sem significado estatístico.

| Modelo    | BIC-BF( $p = 8$ ) |       |       |
|-----------|-------------------|-------|-------|
| Parâmetro | Estimativa        | SE    | p[H0] |
| Intcp.    | -0.017            | 0.006 | 0.004 |
| ei        | -0.959            | 0.083 | 0.000 |
| ae        | 0.537             | 0.064 | 0.000 |
| eg        | 0.843             | 0.163 | 0.000 |
| dg        | -0.310            | 0.133 | 0.020 |
| df        | 0.707             | 0.040 | 0.000 |
| fg        | -0.794            | 0.066 | 0.000 |
| i         | 0.762             | 0.034 | 0.000 |

SE - Erro padrão (estimativa da dispersão do parâmetro); p[H0] - valor de prova

Os resultados da tabela (4.19) não revelam que existe um parâmetro sem significado estatístico.

O modelo encontrado e os parâmetros encontrados encontram-se também

representados na tabela (4.19).

Este modelo apresenta o coeficiente de determinação de 0.8761 enquanto que o desvio padrão do ajuste de  $3.530 \times 10^{-3}$ . Além disso, atentando ainda nos valores da tabela 4.19, verifica-se que o modelo BIC-B( $p = 8$ ) possui todos os parâmetros com significado estatístico.

Assim sendo, o melhor modelo encontrado pelo método BIC-BF é de facto o BIC-BF( $p = 8$ ) cujas estimativas se encontram sistematizadas na tabela 4.19.

#### 4.2.2.5 Interpretação

De acordo com os resultados obtidos na tabela 4.20, para a descrição do teor de dióxido de azoto, vamos proceder a uma análise.

Tabela 4.20: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de dióxido de azoto observado durante o mês de fevereiro de 2017.

| Processo | p | $\sigma_{fit}$ | $R^2$  | AIC     | BIC     | modelo   |
|----------|---|----------------|--------|---------|---------|--|
| AIC-B    | 7 | 0.00378        | 0.8671 | -1831.0 | -1794.8 | 1.03 ai - 1.02 ii + 0.78 df-0.52 dj<br>+ 0.46 dg + 0.44 i+0.15 d           |
| AIC-BF   | 6 | 0.00505        | 0.8223 | -1638.5 | -1606.9 | 1.03 ai +0.49 df -0.47 ei - 0.42 ii<br>+ 0.34 i + 0.01                     |
| BIC-B    | 8 | 0.00353        | 0.8761 | -1876.9 | -1836.3 | -0.95 ei + 0.84 eg - 0.79 fg + 0.76 i<br>+0.70 df + 0.53 ae - 0.31 dg-0.01 |
| BIC-BF   | 8 | 0.00353        | 0.8761 | -1876.9 | -1836.3 | -0.95 ei + 0.84 eg - 0.79 fg + 0.76 i<br>+0.70 df + 0.53 ae - 0.31 dg-0.01 |

a - CO; c - O<sub>3</sub>; d - PM<sub>2.5</sub>; e - PM<sub>10</sub>; f - SO<sub>2</sub>; g - NO;

h - NH<sub>3</sub>; i - NMVOCs; j - PANs; r? - raiz quadrada da variável ?;

Desta tabela é fácil verificar que ambos processos de optimização relacionados com o critério BIC (BIC-B e BIC-BF) conduziram ao mesmo modelo com  $p = 8$  enquanto os modelos encontrados através da optimização com AIC surgem com  $p = 7$  (AIC-B) e  $p = 6$  (AIC-BF).

Através dos indicadores de desempenho verifica-se que os modelos BIC se coadunam mais na descrição dos teores atmosféricos de dióxido de azoto no período em causa.

Focando a atenção no modelo BIC (BIC-B e BIC-BF), verifica-se que os teores mais elevados em dióxido de azoto estão relacionados com:

- maiores valores de NMVOCs (variável “i”);
- maiores valores nas interacções “eg” (PM<sub>10</sub>-NO), “df” (PM<sub>2.5</sub>-SO<sub>2</sub>) e “ae” (CO-PM<sub>10</sub>);
- e menores valores nas interacções “ei” (PM<sub>10</sub>-NMVOCs), “fg” (SO<sub>2</sub>-NO) e “dg” (PM<sub>2.5</sub>-NO).

Atendendo ao que foi anteriormente expresso no Capítulo 1, do ponto vista químico, este modelo relaciona o dióxido de azoto com os NMVOCs - sabe-se que os NMVOCs reagem, por acção da luz solar, com as formas oxidadas de azoto e de enxofre, gerando outras espécies, chamados oxidantes fotoquímicos (espécies de azoto e de enxofre menos oxidadas). Assim, quando os níveis de NMVOCs

umentam, os foto oxidantes gerados devem oxidar as formas menos oxidadas de azoto (NO) gerando maiores teores atmosféricos de dióxido de azoto.

O termo de interacção negativo (PM<sub>10</sub>-NMVOCs) e o termo de interacção positivo (PM<sub>10</sub>-NO) dão uma noção de que as partículas suspensas de maior dimensão (PM<sub>10</sub>) afectam a conversão de NO em NO<sub>2</sub> por via dos NMVOCs - sabe-se que o material particulado de maior dimensão apresenta uma maior opacidade à radiação solar e, deste modo, inibe a fotooxidação de NO a NO<sub>2</sub>.

De modo similar surgem as interacções positiva (PM<sub>2.5</sub>-SO<sub>2</sub>) e negativas (PM<sub>2.5</sub>-NO e SO<sub>2</sub>-NO) que também expressam o mesmo tipo de mecanismo reaccional - o material particulado suspenso de menor dimensão é mais favorável à fotocatalise e conversão de NO em NO<sub>2</sub> por acção da presença de SO<sub>2</sub>.

O termo de interacção positiva (CO-PM<sub>10</sub>) é ainda uma evidência do efeito pernicioso do material particulado de maior dimensão (que impede a fotocatalise oxidativa de NMVOCs bem como o monóxido de carbono como um activo redutor deste processo).

Por uma questão de facilitar a leitura e simplificar a apresentação dos resultados obtidos na modelação das restantes variáveis, no restante trabalho iremos apresentar a tabela de resumo dos melhores modelos para cada uma das restantes variáveis.

### 4.2.3 Ozono

Para facilitar a leitura, vamos passar diretamente dos quatro melhores modelos seleccionados por AIC-B, AIC-BF, BIC-B e BIC-BF. Mais informações podem ser vistas na secção A.1.1 na página 97.

Passamos agora para a análise da escolha do modelo referente ao teor de ozono, ver tabela 4.21.

Tabela 4.21: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de ozono observado durante o mês de fevereiro de 2017.

| Processo | p  | $\sigma_{fit}$ | $R^2$  | AIC     | BIC     | modelo  |
|----------|----|----------------|--------|---------|---------|---|
| AIC-B    | 7  | 0.00613        | 0.7846 | -1507.4 | -1471.3 | -1.74 d + 1.68 ad + 0.99 rj -0.63 aa<br>+ 0.61- 0.43 j - 0.41 a                             |
| AIC-BF   | 7  | 0.00613        | 0.7846 | -1507.4 | -1471.3 | -1.74 d + 1.68 ad + 0.99 rj -0.632 aa<br>+ 0.61- 0.43 j - 0.41 a                            |
| BIC-B    | 10 | 0.01563        | 0.6730 | -875.2  | -825.6  | 2.00 ad - 2.00 d -1.44 aa - 1.37 bf+ 0.91 bb<br>+ 0.81+ 0.47 ff+ 0.39 ai + 0.32 j + 0.162 a |
| BIC-BF   | 10 | 0.01581        | 0.6694 | -867.7  | -818.1  | 2.68 ad- 2.12 d-1.61 aa + 0.83 + 0.74 ai<br>- 0.57 di-0.37 bf + 0.34 j+0.18 ff+0.08 a       |

a - CO; b - NO<sub>2</sub>; d - PM<sub>2.5</sub>; e - PM<sub>10</sub>; f - SO<sub>2</sub>; g - NO;

h - NH<sub>3</sub>; i - NMVOCs; j - PANs; r? - raiz quadrada da variável ?;

Da tabela 4.21 verifica-se que os modelos optimizados através do critério BIC (BIC-B e BIC-BF) são mais complexos e nem por isso capazes de efectuar uma melhor modelação do teor de ozono atmosférico no período em questão.

Verifica-se também que a optimização baseada em AIC foi consistente, independentemente do modo em que ela foi efectuada (AIC-B e AIC-BF).

Atendendo ao melhor modelo encontrado (AIC), maiores níveis de ozono atmosférico estão relacionados com:

- diminuição directa e quadrática de CO;
- diminuição directa de  $PM_{2.5}$ ;
- diminuição de PANs, de uma forma menos linear;
- interacções positivas entre CO- $PM_{2.5}$ .

Em termos de interpretação mecanicista, os níveis de ozono estão relacionados com o efeito da fotoxidação do oxigénio por parte de espécies suspensas oxidadas, presentes, por exemplo nos PANs e  $PM_{2.5}$ .

Já a presença de espécies redutoras, como o monóxido de carbono, têm um efeito directo na redução dos níveis de ozono.

Do ponto vista químico, vemos no modelo, que há uma forte relação negativa, com as variáveis “aa” e “a”, que significam o parâmetro monóxido de carbono, de acordo com [53] esta relação deve-se ao facto de não existir fontes de dióxido de azoto, as espécies estarem isoladas das fontes, e através de processos fotoquímicos reagirem entre si, obtemos assim uma explicação para o nosso modelo, olhando de novo para o modelo, tabela 4.21, encontramos as interacções “ad” e o parâmetro “d” com grande peso no modelo, vamos tentar analisar, no caso “d” referente ao parâmetro  $PM_{2.5}$ , como anteriormente tínhamos visto, as PM em suspensão são um veículo de transporte eficaz de poluentes atmosféricos que se fixam à sua superfície, ou seja, o ozono poderá ficar adsorvido na sua superfície, daí o seu forte relacionamento com esta variável, ao nível de desaparecimento, passando agora para a interacção “ad”, referente aos parâmetros monóxido de azoto com  $PM_{2.5}$ , neste caso é fortemente positiva, analisar modelo tabela 4.21, ou seja, esta a fornecer a espécie ozono, esta interacção transporta a espécie, mas devido as forcas entre as duas, “ad” e ozono, serem mais fracas, devido também, á baixa polarizabilidade da interacção, temperatura, da pressão ou da área da superfície, estes factores podem favorecer a sua menor adsorvidade [52].

#### 4.2.4 $PM_{2.5}$

No que diz respeito à modelação dos teores de partículas suspensas mais finas ( $PM_{2.5}$ ) os modelos seleccionados por AIC-B, AIC-BF, BIC-B e BIC-BF encontram-se resumidos na tabela 4.22. Mais informações podem ser obtidas na secção A.1.2 na página 104.

Tabela 4.22: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de  $PM_{2.5}$  observado durante o mês de fevereiro de 2017.

| Processo | p | $\sigma_{fit}$ | $R^2$  | AIC     | BIC     | modelo  |
|----------|---|----------------|--------|---------|---------|---|
| AIC-B    | 6 | 0.00189        | 0.9226 | -2296.2 | -2264.6 | 0.52 ej+0.25 bi - 0.21 c<br>+ 0.21 + 0.11 gi + 0.07 ii                            |
| AIC-BF   | 6 | 0.00189        | 0.9226 | -2296.2 | -2264.6 | 0.52 ej+0.25 bi - 0.21 c<br>+ 0.21 + 0.11 gi + 0.07 ii                            |
| BIC-B    | 9 | 0.00164        | 0.9334 | -2391.2 | -2346.0 | - 0.53 rc + 0.43 ej+ 0.33+ 0.21 c + 0.19 af<br>+ 0.11 b+0.09 ii+ 0.07 gi- 0.03 bi |
| BIC-BF   | 9 | 0.00164        | 0.9334 | -2391.2 | -2346.0 | - 0.53 rc + 0.43 ej+ 0.33+ 0.21 c + 0.19 af<br>+ 0.11 b+0.09 ii+ 0.07 gi- 0.03 bi |

a - CO; b - NO<sub>2</sub>; c-O<sub>3</sub>; e - PM<sub>10</sub>; f - SO<sub>2</sub>; g - NO;  
h - NH<sub>3</sub>; i - NMVOCs; j - PANs; r? - raiz quadrada da variável ?;

Da tabela 4.22 verifica-se que a optimização conduz a dois tipos de modelos - por um lado o critério AIC (AIC-B e AIC-BF) converge num modelo mais simples ( $p = 6$ ) mas com um pior desempenho em termos de ajuste aos valores reais; por outro lado, o critério BIC (BIC-B e BIC-BF) conduz a um modelo mais complexo ( $p = 9$ ) com o benefício de apresentar melhor desempenho no ajuste aos valores experimentais. O que é interessante observar é que todas as variáveis relevantes no modelo mais simples (AIC-B e AIC-BF) encontram-se presentes no modelo mais completo.

Assumindo que o modelo BIC é o mais fiel na descrição desta variável, em termos de dependências mais relevantes verifica-se que maiores valores de  $PM_{2.5}$  estão relacionados com:

- maiores valores em NO<sub>2</sub>, em O<sub>3</sub> e em NMVOCs (dependência quadrática);
- menores valores em O<sub>3</sub> (dependência não linear);
- maiores interdependências directas com PM<sub>10</sub>-PANs, com CO-SO<sub>2</sub> e com NO-NMVOCs
- menores interdependências directas com NO<sub>2</sub>-NMVOCs.

Analisando o modelo, em termos químicos, podemos verificar a tabela 4.22, constamos que é um modelo complexo, podemos analisar a variável “bi”, correspondente a interacção NO<sub>2</sub> com NMVOCs, espécies são precursoras de  $PM_{2.5}$  como vimos no cap.1 e [54], o que leva a um ponto o  $PM_{2.5}$  do estudo poderá ser um poluente secundário, foi formado a partir destas espécies na atmosfera, passando para a variável “ej”, correspondente a PM<sub>10</sub> e PANs, como referimos anteriormente, estes são considerados bons transportadores de poluentes, adsorvem moléculas [52], devido as suas propriedades, o poderá indiciar devido à forte relação positiva o transporte de  $PM_{2.5}$ .

Concluído o  $PM_{2.5}$  advém de uma grande parte por transporte, por efeito de adsorvidade [55] e por poluição secundária através da relação NO<sub>2</sub> e NMVOCs.

#### 4.2.5 PM<sub>10</sub>

No que diz respeito aos teores de material particulado de maior dimensão (PM<sub>10</sub>), os quatro melhores modelos selecionados por AIC-B, AIC-BF, BIC-



B e BIC-BF encontram-se resumidos na tabela 4.23. Mais informações podem ser vistas A.1.3 na página 110.

Tabela 4.23: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de  $PM_{10}$  observado durante o mês de fevereiro de 2017.

| Processo | p  | $\sigma_{fit}$ | $R^2$  | AIC     | BIC     | modelo   |
|----------|----|----------------|--------|---------|---------|--|
| AIC-B    | 16 | 0.0114         | 0.6709 | -1080.1 | -1003.4 | -2.34 rc+1.54 c- 1.46 dj+1.31 dd -1.23 bi<br>+1.04+ 0.74 bj- 0.65 a+0.64 cd+0.59 ag<br>+0.57 ii-0.54 aa-0.48 g+0.40 ab+0.08 af-0.05 df         |
| AIC-BF   | 17 | 0.0102         | 0.7063 | -1154.7 | -1073.5 | -2.46 rc+ 1.55 c+1.38 dd -1.29 dj+1.12<br>-0.96 bi+0.78 cd+ 0.71 bj +0.69 aa-0.68 a<br>- 0.43 df+0.38 ag+0.37 ii+0.36 af-0.27 g-0.22 h+0.11 ab |
| BIC-B    | 14 | 0.0119         | 0.6555 | -1053.4 | -985.7  | - 2.13 rc+1.45 dd + 1.44 c-1.37 dj-0.97 bi<br>+0.81 +0.57 cd+0.54 bj +0.45 ii+0.44 gi<br>+0.42 ab- 0.40 ag- 0.09 af+ 0.02 df                   |
| BIC-BF   | 14 | 0.0119         | 0.6555 | -1053.4 | -985.7  | - 2.13 rc+1.45 dd + 1.44 c-1.37 dj<br>-0.97 bi+0.81 +0.57 cd+0.54 bj +0.45 ii<br>+0.44 gi+0.42 ab- 0.40 ag- 0.09 af+ 0.02 df                   |

a - CO; b - NO<sub>2</sub>; c-O<sub>3</sub>; d - PM<sub>2.5</sub>; f - SO<sub>2</sub>; g - NO;

h - NH<sub>3</sub>; i - NMVOCs; j - PANs; r? - raiz quadrada da variável ?;

Dos resultados da tabela 4.23 verifica-se que os teores de material particulado de maior dimensão suspenso na atmosfera ( $PM_{10}$ ) tem muitas interações com as restantes variáveis e que estas, por si só, não conseguem descrever com grande fidelidade esta resposta.

Olhando para o conjunto de dados, na tabela 4.23, verificamos dois processos com o mesmo modelo, estes são BIC-B e BIC-BF, em termos de numero de parâmetros são os mais refinados, vamos deixar para o final, passando para os casos dos processos AIC-B e AIC-BF, este ultimo é o que apresenta o melhor desvio do ajuste, é o menor, o coeficiente de determinação, é o mais elevado, em termos dos critérios para o modelo, apresentam valores mínimos o que indica uma boa seleção de parâmetros para a escolha de modelo, mas é neste aspecto que perde contra o conjunto de processos, ver tabela 4.23, apresenta demasiados parâmetros, escolha partirá para AIC-B, BIC-B e BIC-BF, vamos centrar nestes três processos, como vimos anteriormente, o BIC-B e BIC-BF apresentam o mesmo modelo, podemos comparar estes dois em relação ao AIC-B, por uma questão de simplificação na análise, observando os dados da tabela 4.23, em termos de coeficiente de determinação e BIC, diferem um pouco, o que nos leva para análise dos outros termos, AIC e desvio do ajuste são muito próximos, o que indica que em termos de numero de parâmetros, são adequados não perdendo qualidade no ajuste, o que nos leva ao processo BIC-B e BIC-BF, apesar de ter um menor valor de coeficiente de determinação, apresenta um menor numero de parâmetros o que o torna, refinado e com isto o menos complexo do conjunto, que é o que pretendemos também através da modelação, modelos refinados e pouco complexos, mas há um argumento importante contra, coeficiente de determinação diverge muito do desejado, a resposta do modelo não será a ideal,

estes modelos não podem ser selecionados para o teor de  $PM_{10}$  .

Devido ao resultados nos modelos, não poderemos tirar conclusões em termos químicos.

#### 4.2.6 Dióxido de enxofre

Para facilitar a leitura, vamos passar diretamente dos quatro melhores modelos selecionados por AIC-B, AIC-BF, BIC-B e BIC-BF. Mais informações podem ser vistas A.1.4 na página 116

Nesta secção iremos analisar os processos referentes ao teor de dióxido de enxofre, na tabela 4.24 encontramos todos processos e respectivos modelos.

Tabela 4.24: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de dióxido de enxofre observado durante o mês de fevereiro de 2017.

| Processo | p | $\sigma_{fit}$ | $R^2$  | AIC     | BIC     | modelo   |
|----------|---|----------------|--------|---------|---------|--|
| AIC-B    | 7 | 0.00890        | 0.7774 | -1256.9 | -1220.8 | -2.15 ab+1.89 b + 1.04 dj -0.99 dd<br>+ 0.91 ai+0.15 de+0.03       |
| AIC-BF   | 8 | 0.00888        | 0.7780 | -1257.4 | -1216.8 | -2.15 ab+ 1.91 b +1.08 dj-1.06 dd<br>+ 0.98 ai+0.37 d-0.09 a+0.06  |
| BIC-B    | 8 | 0.0103         | 0.7421 | -1156.0 | -1115.3 | -2.29 ab+ 2.05 b +0.91 ae-0.88 de<br>+0.72 dd+ 0.43 bb+0.29 j-0.03 |
| BIC-BF   | 8 | 0.0103         | 0.7421 | -1156.0 | -1115.3 | -2.29 ab+ 2.05 b +0.91 ae-0.88 de<br>+0.72 dd+ 0.43 bb+0.29 j-0.03 |

a - CO; b - NO<sub>2</sub>; c-O<sub>3</sub>; d - PM<sub>2.5</sub>; e - PM<sub>10</sub>; g - NO;

h - NH<sub>3</sub>; i - NMVOCs; j - PANs; r? - raiz quadrada da variável ?;

Ao analisar a tabela 4.24, verificamos dois processos com o mesmo modelo, BIC-B e BIC-BF, vamos optar por uma estratégia diferente, verificando a tabela 4.24, em termos de parâmetros, temos duas escolhas possíveis 7 ou 8, avaliando com os critérios subjacentes ao numero de parâmetros, AIC e BIC, os processos AIC-B e o AIC-BF obtém o menor valor o que corresponde ao melhor modelo em termos de parâmetros sem perder qualidade do ajuste, o que se verifica ao olhar para a tabela 4.24, o AIC-B e o AIC-BF obtém valores muito próximos, mesmo em termos de desvio do ajuste e coeficiente de determinação, o que nos leva a escolher o AIC-BF, por ser o que possui melhor resposta, coeficiente de determinação, para modelar o teor de SO<sub>2</sub>.

Passando para uma pequena análise química do modelo anteriormente selecionado, processo AIC-BF, ver tabela 4.24, constatamos as variáveis com maior peso no modelo, “ab”, “b”, “dj”, “dd” e “ai”, a variável de interação “ab” pode ser considerada um precursor de sulfato (SO<sub>4</sub>), na literatura [56], [57], [58] encontramos este espécie na atmosfera, ao reagir com dióxido de enxofre, a interação fornece átomos de oxigénio na reacção, podem em seguida evoluir para H<sub>2</sub>SO<sub>2</sub>, o ácido sulfúrico, daí aumentar a acidez na atmosfera e surgirem

as chamadas chuvas ácidas, como verificamos no modelo esta interação provoca a diminuição de teor de dióxido de enxofre, assim como a variável, “dd” que corresponde ao segundo grau de  $PM_{2.5}$ , o PM secundário isto é, um poluente secundário é formado a partir desta molécula, dióxido de enxofre, analisar Cap.1, daí a relação negativa no modelo, o PM esta ser formado a partir de dióxido de enxofre, em relação à variável “b”, associada ao dióxido de azoto, através das forcas/interações de Van Der Waals [52] virem absorvidas nestas, assim como as interações “dj” associadas ao  $PM_{2.5}$  e PAN, pode ocorrer o mesmo processo, e correspondente transporte desta molécula, dióxido de enxofre.

#### 4.2.7 Monóxido de azoto

Na tabela 4.25 encontra uma sùmula dos quatro melhores modelos selecionados por AIC-B, AIC-BF, BIC-B e BIC-BF na descrição dos teores atmosféricos em monóxido de azoto. Mais informações podem ser vistas A.1.5 na página 122.

Tabela 4.25: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de monóxido de azoto observado durante o mês de fevereiro de 2017.

| Processo | p  | $\sigma_{fit}$ | $R^2$  | AIC     | BIC     | modelo  |
|----------|----|----------------|--------|---------|---------|---|
| AIC-B    | 8  | 0.00692        | 0.4054 | -1424.2 | -1383.5 | -1.86 bd+1.23 dd-0.97 bb-0.95 bf<br>+0.90 af-0.27 ai+ 0.19 dj+0,01                          |
| AIC-BF   | 8  | 0.00692        | 0.4054 | -1424.2 | -1383.5 | -1.86 bd+1.238 dd-0.97 bb-0.95 bf<br>+0.90 af-0.27 ai+ 0.19 dj+0,01                         |
| BIC-B    | 7  | 0.00700        | 0.3956 | -1415.2 | -1379.1 | -1.71 bd+1.02 dd+0.64 af -0.63 bf<br>+0.58 bb+0.29 dj + 0.01                                |
| BIC-BF   | 11 | 0.00618        | 0.4715 | -1497.3 | -1443.1 | +2.50 de+1.82 bb-1.72 bd-1.45 bf+1.16 af<br>-1.06 ae- 0.60 dd-0.52 be+0.25 dj -0.13 ai+0.01 |

a - CO; b - NO<sub>2</sub>; c-O<sub>3</sub>; d - PM<sub>2.5</sub>; e - PM<sub>10</sub>; f - SO<sub>2</sub>;

h - NH<sub>3</sub>; i - NMVOCs; j - PANs; r? - raiz quadrada da variável ?;

Neste caso, olhando para tabela 4.25, salta-nos à vista o valor referente ao coeficiente de determinação, no conjunto total é um valor muito baixo, inferior a 0.5, o que vai influenciar a resposta do modelo, para minimizar este efeito teremos de optar por uma estratégia ao encontro deste problema, dado que, olhando para a tabela 4.25, os critérios referentes ao numero de parâmetros, AIC e BIC, são muito próximos no conjunto global, e em termos de desvio do ajuste também não ha muita discrepância entre processos, o que nos leva à conclusão, os modelos não exprimem uma relação com o teor de monóxido de azoto.

Os modelos encontrados, não poderão ser examinados em termos químicos, por falta de qualidade dos mesmos.

#### 4.2.8 Amoníaco

Na tabela 4.26 encontram-se reunidos um apontamento dos quatro melhores modelos selecionados por AIC-B, AIC-BF, BIC-B e BIC-BF. Mais informações

podem ser vistas A.1.6 na página 128.

Tabela 4.26: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de amoníaco observado durante o mês de fevereiro de 2017.

| Processo | p  | $\sigma_{fit}$ | $R^2$  | AIC    | BIC    | modelo   |
|----------|----|----------------|--------|--------|--------|--|
| AIC-B    | 7  | 0.0308         | 0.4048 | -421.9 | -385.9 | 9.33 ad-6.32 dd-4.25 aa+2.09 a<br>-0.57 bi -0.42 ab -0.19                                |
| AIC-BF   | 7  | 0.0308         | 0.4048 | -421.9 | -385.9 | 9.33 ad-6.32 dd-4.25 aa+2.09 a<br>-0.57 bi -0.42 ab -0.19                                |
| BIC-B    | 11 | 0.0246         | 0.5275 | -569.0 | -514.9 | 10.87 ad-5.88 aa-5.79 dd-3.87 bi+2.75 bb<br>+ 2.67 ai+2.43 a-1.89 de-0.76 b+0.26 fg-0.18 |
| BIC-BF   | 11 | 0.0246         | 0.5275 | -569.0 | -514.9 | 10.87 ad-5.88 aa-5.79 dd-3.87 bi+2.75 bb<br>+ 2.67 ai+2.43 a-1.89 de-0.76 b+0.26 fg-0.18 |

a - CO; b - NO<sub>2</sub>; c - O<sub>3</sub>; d - PM<sub>2.5</sub>; e - PM<sub>10</sub>; f - SO<sub>2</sub>; g - NO;  
i - NMVOCs; j - PANs; r? - raiz quadrada da variável ?;

Como podemos observar na tabela 4.26, existem dois tipos de modelos, o referente ao AIC-B e AIC-BF, são iguais e o referente ao BIC-B e BIC-BF, também são iguais, como verificamos na tabela 4.26, obtém o mesmo modelo, ficamos assim com dois modelos AICs(AIC-B e AIC-BF) e BICs(BIC-B e BIC-BF), comparando estes dois modelos AICs e BICs, ver tabela 4.26, encontramos um baixo valor de coeficiente de determinação, 0.4 e 0.5 respectivamente, este valor situa-se largamente a baixo do expectável/desejável 0.8-1.0, o que nos leva optar pela estratégia da secção anterior, e seleccionar o modelo que minimiza este factor preponderante na escolha, optamos assim pelo o modelo BICs, ou seja, o BIC-B e BIC-BF, são o que possui maior valor de coeficiente de determinação e baixo valor de desvio do ajuste do modelo, o que evidencia uma melhor qualidade de ajuste, através dos critérios AIC e BIC confirma esta hipótese, ao ser inferior, em relação ao do modelo AICs, o numero de parâmetros é adequado ao tipo de modelo, apesar de não ser o mais refinado, chegamos a um tópico fundamental da modelação, permitir uma melhor interpretação a nível significativo no que se refere à variação da resposta (predita) em função dos preditores/regressores e com este estudo em termos de modelos (baixo valor de coeficiente de determinação), não foi conclusiva a escolha para o teor de amoníaco, não podendo assim tirar qualquer ilação em termos de melhor modelo.

A nível químico, devido à falta de qualidade de modelo, as conclusões referentes às propriedades químicas não serão abordadas neste estudo.

#### 4.2.9 NMVOCs

Nesta secção vamos passar diretamente dos quatro melhores modelos seleccionados por AIC-B, AIC-BF, BIC-B e BIC-BF. Mais informações podem ser vistas A.1.7 na página 134

Passamos agora para a análise da escolha do modelo referente ao teor de NMVOCs, ver tabela 4.27.

Tabela 4.27: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de NMVOCs observado durante o mês de fevereiro de 2017.

| Processo | p | $\sigma_{fit}$ | $R^2$  | AIC     | BIC     | modelo                                   |
|----------|---|----------------|--------|---------|---------|--|
| AIC-B    | 5 | 0.00405        | 0.8790 | -1786.4 | -1759.3 | 0.52 af +0.37 d+0.31 a<br>-0.21 dd- 0.04 |
| AIC-BF   | 5 | 0.00405        | 0.8790 | -1786.4 | -1759.3 | 0.52 af +0.37 d+0.31 a<br>-0.21 dd- 0.04 |
| BIC-B    | 5 | 0.00412        | 0.8770 | -1775.2 | -1748.1 | 0.54 af+0.31 a+0.20 d<br>- 0.02 -0.04 g  |
| BIC-BF   | 5 | 0.00412        | 0.8770 | -1775.2 | -1748.1 | 0.54 af+0.31 a+0.20 d<br>- 0.02 -0.04 g  |

a - CO; b - NO<sub>2</sub>; c - O<sub>3</sub>; d - PM<sub>2.5</sub>; e - PM<sub>10</sub>; f - SO<sub>2</sub>; g - NO;  
h - NH<sub>3</sub>; j - PANs; r? - raiz quadrada da variável ?;

Avaliando os modelos, de acordo com os processos estabelecidos para o teor de monóxido de carbono na tabela 4.27, constatamos que a otimização de modelos através dos critérios AIC-B e AIC-BF são coerentes assim como BIC-B e BIC-BF, já que permitem obter o mesmo modelo.

Verificando a tabela 4.27, ficamos com a categoria AICs (AIC-B e AIC-BF) e categoria BICs (BIC-B e BIC-BF), em termos de parâmetros ambas possui o mesmo numero, cinco, em termos de critérios AIC e BIC há uma ligeira vantagem para a categoria AICs, o que pressupõem um adequado numero de parâmetros para o modelo, obtemos assim uma vantagem, em termos de coeficiente de determinação, constatamos na tabela 4.27, há uma ligeira diferença a favor da categoria AICs, o mesmo sucede para o desvio do ajuste, o que indica que o modelo proposto por esta categoria, exemplifica um melhor ajuste, o que torna os processos AIC-B e AIC-BF os escolhidos para a modelação do teor de NMVOCs.

Se se atender aos modelos obtidos com AICs (AIC-B e AIC-BF) e BICs (BIC-F e BIC-BF), verifica-se que os três primeiros parâmetros são iguais e os mais relevantes evidenciando que as dependências do modelo otimizado com “af”, “d” e com “a” são as mais relevantes, no caso do AICs surge ainda o termo “dd” com algum impacto.

O termo mais relevante relaciona a resposta (níveis de NMVOCs) com “af” (variável que revela a interdependência de CO com SO<sub>2</sub>) - esta interdependência está patente em todos os modelos na tabela 4.27 e assume sempre um valor positivo sugerindo que os níveis de NMVOCs aumentam com os CO e SO<sub>2</sub>-máximos de NMVOCs estarão associados a máximos de “af” (interdependência de CO com SO<sub>2</sub>),

Como evidências ficou que os níveis de NMVOCs atmosférico estão relacionados com:

- efeitos construtivos entre CO-SO<sub>2</sub> acompanhados de um acréscimo simultâneo em CO.

- aumento directo de  $PM_{2.5}$ , dependendo do teor do mesmo isto é o grau associado, em segundo grau ocorre um efeito destrutivo de teor de NMVOCs.

Em termos químicos atmosféricos, os NMVOCs participam em reacções fotoquímicas complexas, tais como a formação de ozono e PANs [59], no modelo BICs surge uma evidencia a variável “g” associada ao produto de formação de ozono ver Cap.1, o que pode de certo modo, indiciar a variável “af” (interdependência de CO com  $SO_2$ ) e a variável “a”(CO), serem resultados de produtos de reacções complexas na formação de PANs, ozono e de outros oxidantes tóxicos, podemos observar na figura 4.9 no período entre as 9h e as 19h, diminuição do teor de NMVOCs a contrastar com a figura 4.10 (PANs) e da figura 4.3 ( $O_3$ ), um aumento dos correspondentes teores neste período, apesar de não aparecer nenhuma relação directa, isto é, variável “c” e “j” nos modelos, de acordo com [59], o que poderão descrever produtos de reacções de outros oxidantes que participam no processo.

O segundo termo mais relevante “d” embora no modelo BICs, corresponde ao terceiro, expressa a dependência de NMVOCs com  $PM_{2.5}$ , ou seja em termos químicos da atmosférica, vai de acordo com estabelecido anteriormente, o material particulado funciona como transportador de poluentes por adsorvidade[55], o grau de adsorção depende da temperatura, da pressão e da área da superfície, é necessário um estudo com mais dados, em termos de períodos anuais para poder tirar melhores ilações. Associado ao modelo AICs surge o termo “dd”(dependência de segundo grau de  $PM_{2.5}$ ), o que evidencia que dado a correspondência negativa no modelo, aumento considerável (dependência de segundo grau) de  $PM_{2.5}$  poderá estar a ser produzido através de NMVOCs, neste caso como poluente secundário, devido ao abaixamento do teor de NMVOCs, como referido no Cap.1.

#### 4.2.10 PANs

Para facilitar a leitura, nesta secção vamos passar directamente dos quatro melhores modelos seleccionados por AIC-B, AIC-BF, BIC-B e BIC-BF. Mais informações podem ser vistas A.1.8 na página 140

A descrição dos processos, relativos aos modelos de teor de PANs, estão presentes na tabela 4.28.

Tabela 4.28: Resumo dos melhores modelos obtidos para a descrição dos valores registados para o teor de PANs observado durante o mês de fevereiro de 2017.

| Processo | p  | $\sigma_{fit}$ | $R^2$  | AIC    | BIC    | modelo  |
|----------|----|----------------|--------|--------|--------|---|
| AIC-B    | 11 | 0.0212         | 0.5075 | -666.8 | -612.6 | -3.45 bi+3.35 bd+ 2.09 ab+1.72i -1.68 b<br>-1.59 de+0.91 g-0.71 ag+0.34 ae-0.22 be+0,17 |
| AIC-BF   | 11 | 0.0212         | 0.5075 | -666.8 | -612.6 | -3.45 bi+3.35 bd+ 2.09 ab+1.72i -1.68 b<br>-1.59 de+0.91 g-0.71 ag+0.34 ae-0.22 be+0,17 |
| BIC-B    | 7  | 0.0257         | 0.4015 | -543.7 | -507.6 | 4.05 ab-3.65 bi-2.43 b +1.98 i<br>+1.59 be-1.39 ae + 0.25                               |
| BIC-BF   | 7  | 0.0257         | 0.4015 | -543.7 | -507.6 | 4.05 ab-3.65 bi-2.43 b +1.98 i<br>+1.59 be-1.39 ae + 0.25                               |

a - CO; b - NO<sub>2</sub>; c - O<sub>3</sub>; d - PM<sub>2.5</sub>; e - PM<sub>10</sub>; f - SO<sub>2</sub>; g - NO;  
h - NH<sub>3</sub>; i - NMVOCs; r? - raiz quadrada da variável ?;

Como podemos observar na tabela 4.28, existem dois tipos de modelos, o referente ao AIC-B e AIC-BF, são iguais e o referente ao BIC-B e BIC-BF, também são iguais, como verificamos na tabela 4.28, obtém o mesmo modelo, ficamos assim com dois modelos AICs(AIC-B e AIC-BF) e BICs(BIC-B e BIC-BF), comparando estes dois modelos AICs e BICs, ver tabela 4.28, encontramos um baixo valor de coeficiente de determinação, 0.5 e 0.4 respectivamente, este valor situa-se largamente a baixo do expectável/desejável 0.8-1.0, o que nos leva optar pela estratégia das secções anteriores, e seleccionar o modelo que minimiza este factor preponderante na escolha, optamos assim pelo o modelo AICs, ou seja, o AIC-B e AIC-BF, são o que possui maior valor de coeficiente de determinação e baixo valor de desvio do ajuste do modelo, o que evidencia uma melhor qualidade de ajuste, através dos critérios AIC e BIC confirma esta hipótese, ao ser inferior, em relação ao do modelo AICs, o numero de parâmetros é adequado ao tipo de modelo, apesar de não ser o mais refinado, um modelo mais complexo, chegamos a um tópico fundamental da modelação, permitir uma melhor interpretação a nível significativo no que se refere à variação da resposta (predita) em função dos preditores/regressores, a escolha para o teor de PANs , com este estudo não foi conclusiva em termos de modelos (baixo valor de coeficiente de determinação).

Devido à falta de qualidade dos modelos, não permite retirar conclusões em termos químicos do teor de PANs.

## Capítulo 5

# Conclusões

Este estudo procurou evidenciar que a qualidade do ar pode ser avaliada em termos da sua caracterização química, diagnosticada através da análise espectral obtida com imagens de satélite.

Através da modelação multivariada e suas estratégias de optimização e verificação de modelos, é possível obter informação quantificada no que diz respeito à influência e interdependência cruzada das diferentes vertentes de caracterização química da qualidade do ar.

Este estudo, embora ainda algo incipiente, permitiu-nos avaliar o efeito contributivo e intrincado de diversas variáveis em redor de um determinado elemento de diagnóstico da qualidade do ar, evidenciando que existem uma grande quantidade de processos simultâneos interrelacionados.

Este trabalho evidenciou uma boa descrição (modelação) dos teores de  $PM_{2.5}$ , em termos de desempenho, coeficiente de determinação 0.9334, perto do desejável 1.0, um excelente valor, a avaliar pelo desvio do ajuste 0.00164, muito baixo, através de referidos critérios AIC e BIC demonstra a qualidade do modelo, um caso bastante promissor. Podemos também referir os casos de  $NO_2$ , CO e NMVOCs, modelos que em termos de qualidade de ajuste se assimilam, termo do coeficiente de determinação variam entre 0.84 e 0.88, desvio do ajuste entre 0.0035 e 0.0040, o que evidencia uma boa qualidade do ajuste, uma resposta precisa do modelo com os parâmetros a variar entre 5 e 8 para a descrição destes modelos, o que demonstra a grande potencialidade do método imbuído na elaboração do estudo.

Para concluir, através deste trabalho podemos interpretar em termos químicos, as reacções e interações moleculares, através de ferramentas estatísticas e tecnologias avançadas, de relevar que podemos alcançar domínios nunca antes atingidos, de referir que este estudo, devido às limitações quanto aos dados disponíveis, por se tratarem de novas tecnologias ainda possuem algumas limitações, ainda estão em fase de teste, só permitiu uma recolha de dados por um período de um mês, mas apesar de reduzida a informação conseguimos obter alguns modelos com significado em termos estatísticos, o que demonstra a potencialidade do método, para ilações mais claras em termos químicos, teremos de recolher dados por um período muito superior de forma a avaliar a robustez dos nossos modelos, teremos também de esperar novos avanços em termos tecnológicos, espectrômetros com maior poder de resolução em satélites, este ano como referi o lançamento do sentinel-5P, vai ser mais um passo no que toca



a recolha de dados de poluição através destas poderosas ferramentas, mas não ficamos por aqui, uma nova geração de satélites estará a caminho em 2020,[60] os sentinel-5 e sentinel-6 focados a tratar informação relativa á qualidade do ar, isto tudo para referir alguns dos avanços significativos na matéria.

# Bibliografia

- [1] Daniel Vallero. *Fundamentals of air pollution*. Academic press, 2014.
- [2] Cristina Guerreiro. Air quality in europe: 2013 report. 2013.
- [3] Cristina B.B. Guerreiro, Valentin Foltescu, and Frank de Leeuw. Air quality status and trends in europe. *Atmospheric Environment*, 98:376 – 384, 2014.
- [4] World Health Organization (WHO et al. Air quality guidelines for europe. 2009.
- [5] Han Shi, Yutao Wang, Jianmin Chen, and Donald Huisingh. Preventing smog crises in china and globally. *Journal of Cleaner Production*, 112:1261 – 1271, 2016. Preventing Smog Crises.
- [6] Tunde Sule. The influence of primary air pollutants on human health related risk. 3:147–156, 08 2013.
- [7] Michelle L Bell, Devra L Davis, and Tony Fletcher. A retrospective assessment of mortality from the london smog episode of 1952: the role of influenza and pollution. *Environmental health perspectives*, 112(1):6, 2004.
- [8] Elliot Wagland. 11 incredible pictures from the great smog of 1952 @ONLINE, 2013.
- [9] Philippe Leuenberger, Joel Schwartz, Ursula Ackermann-Liebrich, Kurt Blaser, Gianfranco Bolognini, Jean-Pierre Bongard, O Brandli, P Braun, C Bron, and M Brutsche. Passive smoking exposure in adults and chronic respiratory symptoms (sapaldia study). swiss study on air pollution and lung diseases in adults, sapaldia team. *American journal of respiratory and critical care medicine*, 150(5):1222–1228, 1994.
- [10] Denise Felber Dietrich, Christian Schindler, Joel Schwartz, Jean-Claude Barthélémy, Jean-Marie Tschopp, Frédéric Roche, Arnold von Eckardstein, Otto Brändli, Philipppe Leuenberger, Diane R Gold, et al. Heart rate variability in an ageing population and its association with lifestyle and cardiovascular risk factors: results of the sapaldia study. *Europace*, 8(7):521–529, 2006.
- [11] Brunello Wüthrich, Peter Schmid-Grendelmeier, Christian Schindler, Me-dea Imboden, Andreas Bircher, Elisabeth Zemp, Nicole Probst-Hensch, Sapaldia Team, et al. Prevalence of atopy and respiratory allergic diseases in the elderly sapaldia population. *International archives of allergy and immunology*, 162(2):143–148, 2013.

- [12] European Respiratory Society. *Qualite de lair et sante*, 2010.
- [13] Jianjun He, Sunling Gong, Ye Yu, Lijuan Yu, Lin Wu, Hongjun Mao, Congbo Song, Suping Zhao, Hongli Liu, Xiaoyu Li, and Ruipeng Li. Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major chinese cities. *Environmental Pollution*, 223:484–496, 2017.
- [14] Chris A McLinden, Vitali Fioletov, Mark W Shephard, Nick Krotkov, Can Li, Randall V Martin, Michael D Moran, and Joanna Joiner. Space-based detection of missing sulfur dioxide sources of global air pollution. *Nature Geoscience*, 9(7):496–500, 2016.
- [15] Joel S Levine, Tommy R Augustsson, Iris C Andersont, James M Hoell, and Dana A Brewer. Tropospheric sources of nox: lightning and biology. *Atmospheric Environment (1967)*, 18(9):1797–1804, 1984.
- [16] David Fowler, Chris Flechard, Ute Skiba, Mhairi Coyle, and J Neil Cape. The atmospheric budget of oxidized nitrogen and its role in ozone formation and deposition. *The New Phytologist*, 139(1):11–23, 1998.
- [17] John H Seinfeld and Spyros N Pandis. *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons, 2016.
- [18] Gregor Kiesewetter, Wolfgang Schoepp, Chris Heyes, and Markus Amann. Modelling pm 2.5 impact indicators in europe: Health effects and legal compliance. *Environmental Modelling & Software*, 74:201–211, 2015.
- [19] Ole Raaschou-Nielsen, Zorana J Andersen, Rob Beelen, Evangelia Samoli, Massimo Stafoggia, Gudrun Weinmayr, Barbara Hoffmann, Paul Fischer, Mark J Nieuwenhuijsen, Bert Brunekreef, et al. Air pollution and lung cancer incidence in 17 european cohorts: prospective analyses from the european study of cohorts for air pollution effects (escape). *The lancet oncology*, 14(9):813–822, 2013.
- [20] Shagoon Tabin. *Global Warming: The Effect Of Ozone Depletion*. APH Publishing, 2008.
- [21] Joe L Lane. Stratospheric ozone depletion. In *Life Cycle Impact Assessment*, pages 51–73. Springer, 2015.
- [22] Sagar V Krupa and William J Manning. Atmospheric ozone: formation and effects on vegetation. *Environmental Pollution*, 50(1-2):101–137, 1988.
- [23] N Grossblatt. *Ozone and other photochemical oxidants*. 1977.
- [24] Claire E Reeves, Stuart A Penkett, Stephane Bauguitte, Kathy S Law, Mathew J Evans, Brian J Bandy, Paul S Monks, Gavin D Edwards, Gavin Phillips, Hannah Barjat, et al. Potential for photochemical ozone formation in the troposphere over the north atlantic as derived from aircraft observations during acsoe. *Journal of Geophysical Research: Atmospheres*, 107(D23), 2002.

- [25] BW LaFranchi, GM Wolfe, JA Thornton, SA Harrold, EC Browne, KE Min, PJ Wooldridge, JB Gilman, WC Kuster, PD Goldan, et al. Closing the peroxy acetyl nitrate budget: observations of acyl peroxy nitrates (pan, ppn, and mpan) during bearpex 2007. *Atmospheric Chemistry and Physics*, 9(19):7623–7641, 2009.
- [26] Lars Olof Björn. Stratospheric ozone, ultraviolet radiation, and cryptogams. *Biological Conservation*, 135(3):326 – 333, 2007. The Conservation Ecology of Cryptogams.
- [27] S. Armenta and M. de la Guardia. Pollutants and air pollution. *Comprehensive Analytical Chemistry*, 73:27 – 44, 2016. The Quality of Air.
- [28] Bhola R Gurjar, Luisa T Molina, and C Shekhar P Ojha. *Air pollution: health and environmental impacts*. CRC press, 2010.
- [29] Paolo Zannetti. *Air pollution modeling: theories, computational methods and available software*. Springer Science & Business Media, 2013.
- [30] ESA website. Copernicus, 2017.
- [31] ESA website. Sentinel-5p, 2014.
- [32] Amirhossein Adami, Mahdi Mortazavi, Mehran Nosratollahi, Mohammadreza Taheri, and Jalal Sajadi. Multidisciplinary design optimization and analysis of hydrazine monopropellant propulsion system. *International Journal of Aerospace Engineering*, 2015, 2015.
- [33] ESA website. Sentinel-5p, 2017.
- [34] Britannica website. Electromagnetic spectrum, 2017.
- [35] David G. Streets, Timothy Canty, Gregory R. Carmichael, Benjamin de Foy, Russell R. Dickerson, Bryan N. Duncan, David P. Edwards, John A. Haynes, Daven K. Henze, Marc R. Houyoux, Daniel J. Jacob, Nikolay A. Krotkov, Lok N. Lamsal, Yang Liu, Zifeng Lu, Randall V. Martin, Gabriele G. Pfister, Robert W. Pinder, Ross J. Salawitch, and Kevin J. Wecht. Emissions estimation from satellite retrievals: A review of current capability. *Atmospheric Environment*, 77:1011 – 1042, 2013.
- [36] Paul Ingmann, Ben Veihelmann, Jorg Langen, and Yasjka Meijer. Mission objectives of the atmospheric composition related sentinels s5p, s4, and s5. In *ESA Living Planet Symposium*, volume 722, page 138, 2013.
- [37] Sapcelayer website. Project soul, 2017.
- [38] James N Miller and Jane Charlotte Miller. *Statistics and chemometrics for analytical chemistry*. Pearson Education, 2005.
- [39] Douglas C Montgomery and George C Runger. *Applied statistics and probability for engineers*. John Wiley and Sons, 2010.
- [40] Otaviano Helene. *Metodos dos Minimos Quadrados*. Editora Livraria da Física, 2006.

- [41] Nico JD Nagelkerke. A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692, 1991.
- [42] Htrotugu Akaike. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60(2):255–265, 1973.
- [43] Kenneth P Burnham and David R Anderson. Multimodel inference: understanding aic and bic in model selection. *Sociological methods and research*, 33(2):261–304, 2004.
- [44] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.
- [45] Clifford M Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.
- [46] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.
- [47] Yuhong Yang. Can the strengths of aic and bic be shared? a conflict between model identification and regression estimation. *Biometrika*, 92(4):937–950, 2005.
- [48] Zhongheng Zhang. Variable selection with stepwise and best subset approaches. *Annals of translational medicine*, 4(7), 2016.
- [49] Lars Moseholm, Jeff Silva, and Timothy Larson. Forecasting carbon monoxide concentrations near a sheltered intersection using video traffic surveillance and neural networks. *Transportation Research Part D: Transport and Environment*, 1(1):15–28, 1996.
- [50] Bernardo Beckerman, Michael Jerrett, Jeffrey R Brook, Dave K Verma, Muhammad A Arain, and Murray M Finkelstein. Correlation of nitrogen dioxide with other traffic pollutants near a major expressway. *Atmospheric Environment*, 42(2):275–290, 2008.
- [51] WL Meerts, FH De Leeuw, and A Dymanus. Electric and magnetic properties of carbon monoxide by molecular-beam electric-resonance spectroscopy. *Chemical Physics*, 22(2):319–324, 1977.
- [52] Lucia Ferrari, Josef Kaufmann, Frank Winnefeld, and Johann Plank. Interaction of cement model systems with superplasticizers investigated by atomic force microscopy, zeta potential, and adsorption measurements. *Journal of Colloid and Interface Science*, 347(1):15–24, 2010.
- [53] Hollowayl Parrish, Trainer. Relationships between ozone and carbon monoxide at surface sites in the north atlantic region. *Geophysical Research Letters*, 103(11):13357–13376, 1998.
- [54] Global atmospheric pollution forum website. Copernicus, 2017.
- [55] Jack George Calvert. Glossary of atmospheric chemistry terms (recommendations 1990). *Pure and applied chemistry*, 62(11):2167–2219, 1990.

- [56] ES Saltzman, DL Savoie, RG Zika, and JM Prospero. Methane sulfonic acid in the marine atmosphere. *Journal of Geophysical Research: Oceans*, 88(C15):10897–10902, 1983.
- [57] Noriyuki Tanaka, Danny M Rye, Yitian Xiao, and Antonio C Lasaga. Use of stable sulfur isotope systematics for evaluating oxidation reaction pathways and in-cloud-scavenging of sulfur dioxide in the atmosphere. *Geophysical Research Letters*, 21(14):1519–1522, 1994.
- [58] Hong He, Yuesi Wang, Qingxin Ma, Jinzhu Ma, Biwu Chu, Dongsheng Ji, Guiqian Tang, Chang Liu, Hongxing Zhang, and Jiming Hao. Mineral dust and nox promote the conversion of so2 to sulfate in heavy pollution days. *Scientific reports*, 4:4172, 2014.
- [59] Pratap Kumar Padhy and C.K. Varshney. Total non-methane volatile organic compounds (tnmvoc) in the atmosphere of delhi. *Atmospheric Environment*, 34(4):577 – 584, 2000.
- [60] Josef Aschbacher. Evolution of the copernicus space component: preparing for tomorrow. In *EGU General Assembly Conference Abstracts*, volume 18, page 17824, 2016.

# Apêndice A

## Material suplementar

### A.1 Análise multivariada

Nesta secção encontram-se reunidas as informações mais relevantes acerca dos modelos multivariados mais relevantes explorados para as restantes variáveis relacionadas com a qualidade do ar.

#### A.1.1 Ozono

Nas quatro secções seguintes apresenta-se um resumo dos melhores modelos obtidos para o caso do estudo da modelação do teor de ozono no mês de fevereiro de 2017.

##### A.1.1.1 AIC-B

Utilizando o processo de simplificação do modelo com base no critério de Akaike simplificativo (AIC-B) obtiveram-se as estimativas .

Tabela A.1: Evolução dos indicadores de desempenho do modelo AIC-B em função do subconjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC    | R <sup>2</sup> |
|----|----|--|----------------|---------|--------|----------------|
| 20 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi+rb+bf+aj+ah | 0.01201        | -1043.1 | -948.3 | 0.7528         |
| 19 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi+rb+bf+aj    | 0.01202        | -1043.3 | -953.1 | 0.7521         |
| 18 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi+rb+bf       | 0.01250        | -1017.7 | -932.0 | 0.7417         |
| 17 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi+rb          | 0.01249        | -1019.5 | -938.4 | 0.7417         |
| 16 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi             | 0.01248        | -1020.8 | -944.2 | 0.7414         |
| 15 | 12 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii                | 0.01259        | -1016.1 | -943.9 | 0.7388         |
| 14 | 12 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de                   | 0.01291        | -1000.1 | -932.4 | 0.7317         |
| 13 | 11 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb                      | 0.01313        | -989.9  | -926.7 | 0.7268         |
| 12 | 9  | aa+a+ad+j+d+rj+di+b+ai+ra+ff                         | 0.01375        | -959.6  | -901.0 | 0.7133         |
| 11 | 7  | aa+a+ad+j+d+rj+di+b+ai+ra                            | 0.01422        | -938.1  | -883.9 | 0.7031         |
| 10 | 9  | aa+a+ad+j+d+rj+di+b+ai                               | 0.01429        | -935.7  | -886.1 | 0.7012         |
| 9  | 8  | aa+a+ad+j+d+rj+di+b                                  | 0.01451        | -926.4  | -881.3 | 0.6961         |
| 8  | 7  | aa+a+ad+j+d+rj+di                                    | 0.01510        | -900.7  | -860.1 | 0.6833         |
| 7  | 7  | aa+a+ad+j+d+rj                                       | 0.01511        | -901.1  | -865.0 | 0.6826         |
| 6  | 5  | aa+a+ad+j+d  | 0.01616        | -857.1  | -825.5 | 0.6601         |
| 5  | 5  | aa+a+ad+j  | 0.01942        | -734.7  | -707.6 | 0.5909         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

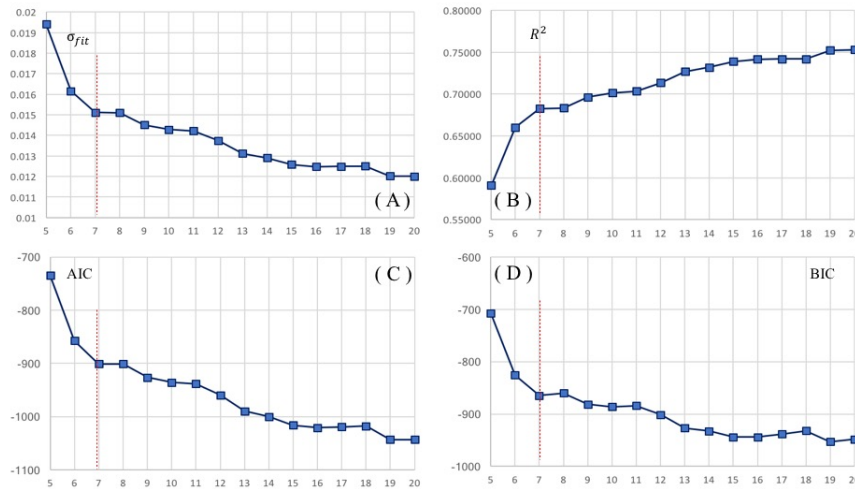


Figura A.1: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o subconjunto de parâmetros refinados através do critério AIC-Backward.



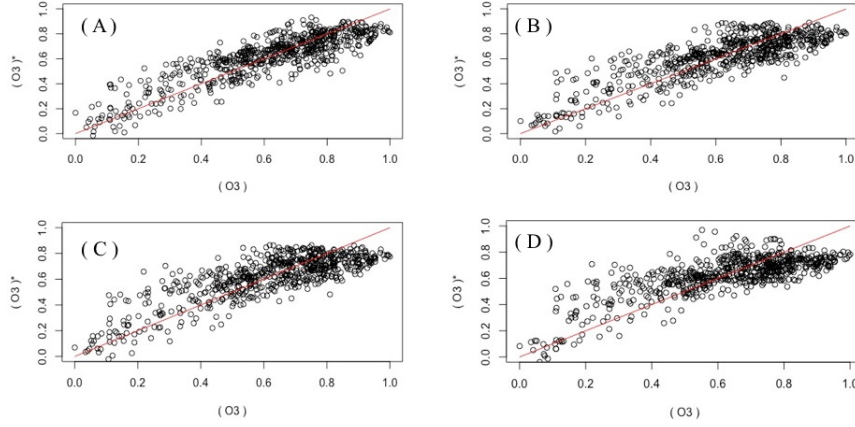


Figura A.2: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 7$ , (C) modelo deficitário com  $p = 6$  e (D) modelo extremo com  $p = 5$ .

**A.1.1.2 AIC-BF**

Tabela A.2: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC    | R <sup>2</sup> |
|----|----|--|----------------|---------|--------|----------------|
| 20 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi+rb+bf+aj+ah | 0.01201        | -1043.1 | -948.3 | 0.7528         |
| 19 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi+rb+bf+aj    | 0.01202        | -1043.3 | -953.1 | 0.7521         |
| 18 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi+rb+bf       | 0.01250        | -1017.7 | -932.0 | 0.7417         |
| 17 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi+rb          | 0.01249        | -1019.5 | -938.4 | 0.7417         |
| 16 | 13 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii+fi             | 0.01248        | -1020.8 | -944.2 | 0.7414         |
| 15 | 12 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de+ii                | 0.01259        | -1016.1 | -943.9 | 0.7388         |
| 14 | 12 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb+de                   | 0.01291        | -1000.1 | -932.4 | 0.7317         |
| 13 | 11 | aa+a+ad+j+d+rj+di+b+ai+ra+ff+bb                      | 0.01313        | -989.9  | -926.7 | 0.7268         |
| 12 | 9  | aa+a+ad+j+d+rj+di+b+ai+ra+ff                         | 0.01375        | -959.6  | -901.0 | 0.7133         |
| 11 | 7  | aa+a+ad+j+d+rj+di+b+ai+ra                            | 0.01422        | -938.1  | -883.9 | 0.7031         |
| 10 | 9  | aa+a+ad+j+d+rj+di+b+ai                               | 0.01429        | -935.7  | -886.1 | 0.7012         |
| 9  | 8  | aa+a+ad+j+d+rj+di+b                                  | 0.01451        | -926.4  | -881.3 | 0.6961         |
| 8  | 7  | aa+a+ad+j+d+rj+di                                    | 0.01510        | -900.7  | -860.1 | 0.6833         |
| 7  | 7  | aa+a+ad+j+d+rj                                       | 0.01511        | -901.1  | -865.0 | 0.6826         |
| 6  | 5  | aa+a+ad+j+d  | 0.01616        | -857.1  | -825.5 | 0.6601         |
| 5  | 5  | aa+a+ad+j  | 0.01942        | -734.7  | -707.6 | 0.5909         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

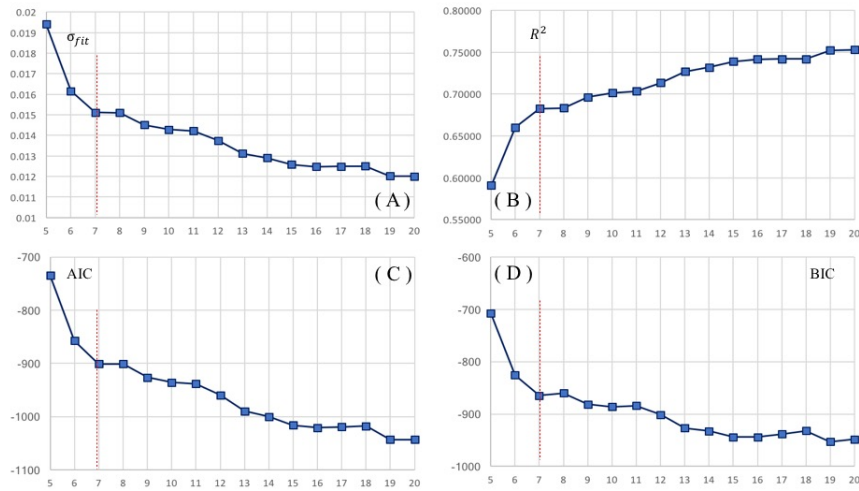


Figura A.3: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

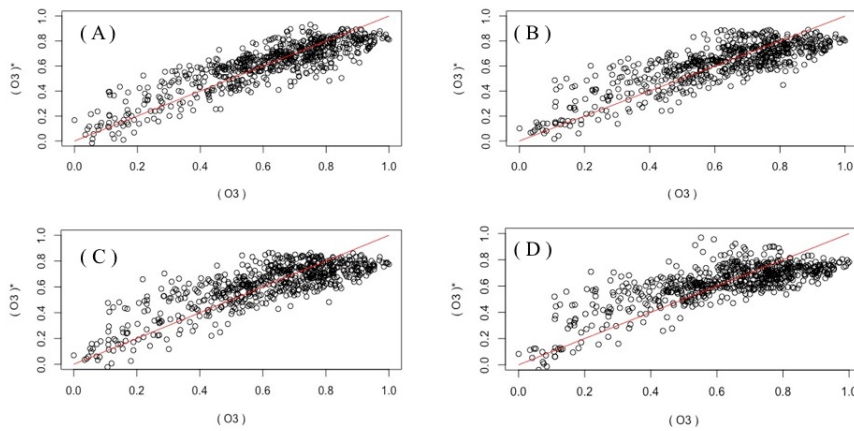


Figura A.4: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 7$ , (C) modelo deficitário com  $p = 6$  e (D) modelo extremo com  $p = 5$ .

A.1.1.3 BIC-B

Tabela A.3: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC    | R <sup>2</sup> |
|----|----|--|----------------|---------|--------|----------------|
| 20 | 12 | aa+ad+a+d+bf+ai+ff+bb+j+di+rj+b+ra+fg+ej+fj+rb+gg+ah | 0.01155        | -1069.1 | -974.4 | 0.7622         |
| 19 | 12 | aa+ad+a+d+bf+ai+ff+bb+j+di+rj+b+ra+fg+ej+fj+rb+gg    | 0.01169        | -1061.9 | -971.7 | 0.7589         |
| 18 | 13 | aa+ad+a+d+bf+ai+ff+bb+j+di+rj+b+ra+fg+ej+fj+rb       | 0.01168        | -1063.6 | -977.9 | 0.7588         |
| 17 | 14 | aa+ad+a+d+bf+ai+ff+bb+j+di+rj+b+ra+fg+ej+fj          | 0.01167        | -1065.2 | -984.0 | 0.7586         |
| 16 | 11 | aa+ad+a+d+bf+ai+ff+bb+j+di+rj+b+ra+fg+ej             | 0.01179        | -1058.9 | -982.2 | 0.7556         |
| 15 | 11 | aa+ad+a+d+bf+ai+ff+bb+j+di+rj+b+ra+fg                | 0.01313        | -987.9  | -915.7 | 0.7276         |
| 14 | 10 | aa+ad+a+d+bf+ai+ff+bb+j+di+rj+b+ra                   | 0.01314        | -988.2  | -920.5 | 0.7269         |
| 13 | 11 | aa+ad+a+d+bf+ai+ff+bb+j+di+rj+b                      | 0.01317        | -987.9  | -924.8 | 0.7260         |
| 12 | 9  | aa+ad+a+d+bf+ai+ff+bb+j+di+rj                        | 0.01453        | -922.5  | -863.9 | 0.6971         |
| 11 | 9  | aa+ad+a+d+bf+ai+ff+bb+j+di                           | 0.01564        | -874.3  | -820.2 | 0.6736         |
| 10 | 9  | aa+ad+a+d+bf+ai+ff+bb+j                              | 0.01564        | -875.2  | -825.6 | 0.6730         |
| 9  | 8  | aa+ad+a+d+bf+ai+ff+bb                                | 0.01828        | -771.2  | -726.1 | 0.6172         |
| 8  | 7  | aa+ad+a+d+bf+ai+ff                                   | 0.01910        | -743.0  | -702.4 | 0.5996         |
| 7  | 7  | aa+ad+a+d+bf+ai                                      | 0.01981        | -719.4  | -683.3 | 0.5840         |
| 6  | 6  | aa+ad+a+d+bf   | 0.01998        | -714.7  | -683.1 | 0.5798         |
| 5  | 5  | aa+ad+a+d  | 0.02042        | -700.9  | -673.9 | 0.5699         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

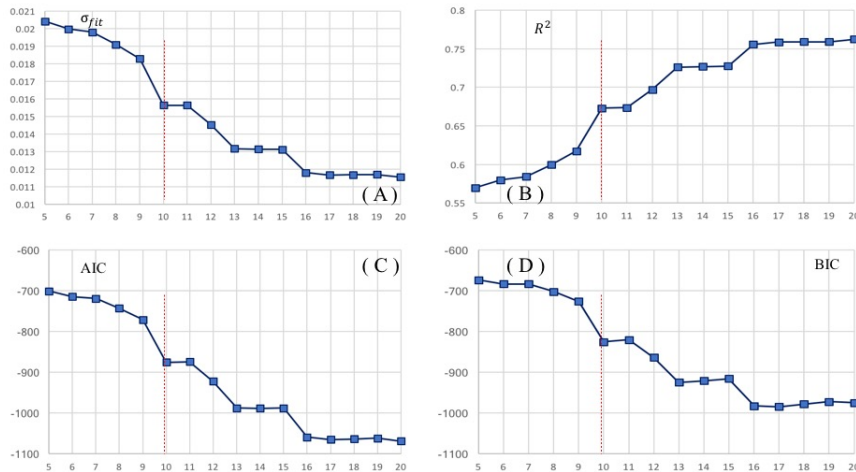


Figura A.5: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward.

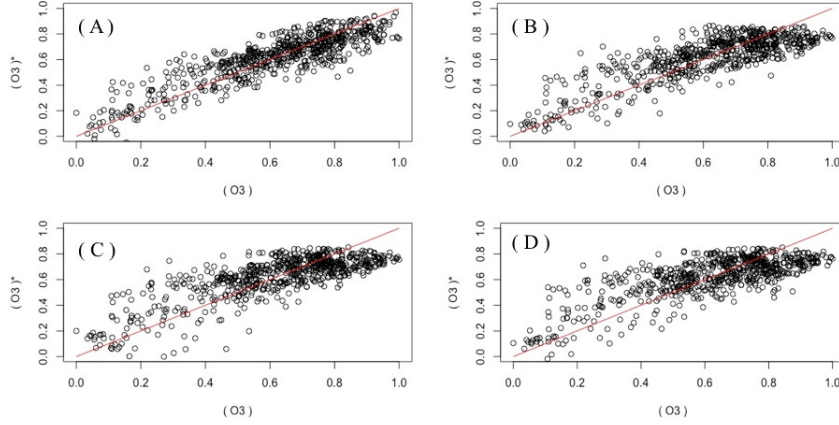


Figura A.6: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 10$ , (C) modelo deficitário com  $p = 9$  e (D) modelo extremo com  $p = 5$ .

**A.1.1.4 BIC-BF**

Tabela A.4: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC    | R <sup>2</sup> |
|----|----|--|----------------|---------|--------|----------------|
| 20 | 12 | aa+a+d+ad+ff+di+ai+bf+j+bb+b+rj+de+ii+ra+ei+fg+fi+rb | 0.01253        | -1014.5 | -919.8 | 0.7420         |
| 19 | 12 | aa+a+d+ad+ff+di+ai+bf+j+bb+b+rj+de+ii+ra+ei+fg+fi    | 0.01252        | -1016.0 | -925.8 | 0.7418         |
| 18 | 11 | aa+a+d+ad+ff+di+ai+bf+j+bb+b+rj+de+ii+ra+ei+fg       | 0.01259        | -1013.0 | -927.3 | 0.7399         |
| 17 | 11 | aa+a+d+ad+ff+di+ai+bf+j+bb+b+rj+de+ii+ra+ei          | 0.01263        | -1012.1 | -931.0 | 0.7388         |
| 16 | 12 | aa+a+d+ad+ff+di+ai+bf+j+bb+b+rj+de+ii+ra             | 0.01261        | -1014.1 | -937.5 | 0.7388         |
| 15 | 13 | aa+a+d+ad+ff+di+ai+bf+j+bb+b+rj+de+ii                | 0.01262        | -1014.6 | -942.4 | 0.7382         |
| 14 | 12 | aa+a+d+ad+ff+di+ai+bf+j+bb+b+rj+de                   | 0.01296        | -997.5  | -929.9 | 0.7307         |
| 13 | 11 | aa+a+d+ad+ff+di+ai+bf+j+bb+b+rj                      | 0.01317        | -987.9  | -924.8 | 0.7260         |
| 12 | 10 | aa+a+d+ad+ff+di+ai+bf+j+bb+b                         | 0.01366        | -964.2  | -905.5 | 0.7153         |
| 11 | 9  | aa+a+d+ad+ff+di+ai+bf+j+bb                           | 0.01564        | -874.3  | -820.2 | 0.6736         |
| 10 | 8  | aa+a+d+ad+ff+di+ai+bf+j                              | 0.01581        | -867.7  | -818.1 | 0.6694         |
| 9  | 7  | aa+a+d+ad+ff+di+ai+bf                                | 0.01907        | -742.8  | -697.7 | 0.6007         |
| 8  | 5  | aa+a+d+ad+ff+di+ai                                   | 0.02041        | -698.4  | -657.8 | 0.5721         |
| 7  | 5  | aa+a+d+ad+ff+di                                      | 0.02044        | -698.3  | -662.2 | 0.5708         |
| 6  | 5  | aa+a+d+ad+ff   | 0.02041        | -700.1  | -668.5 | 0.5706         |
| 5  | 5  | aa+a+d+ad  | 0.02042        | -700.9  | -673.9 | 0.5699         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

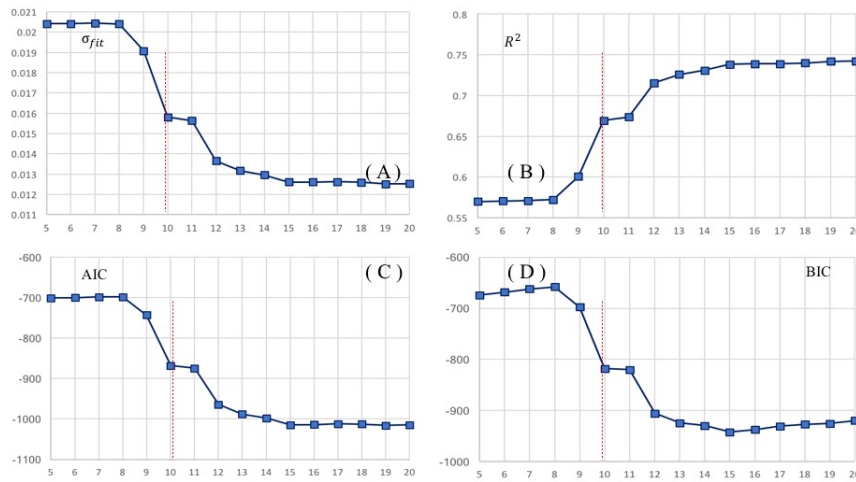


Figura A.7: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward-Forward.

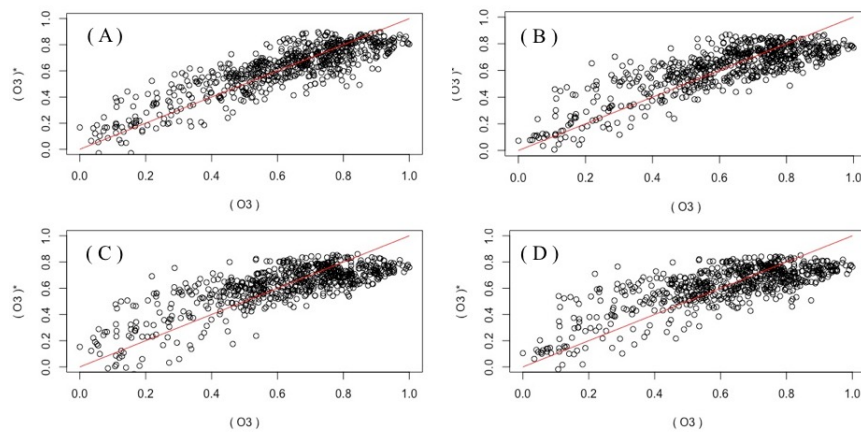


Figura A.8: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 10$ , (C) modelo deficitário com  $p = 9$  e (D) modelo extremo com  $p = 5$ .

**A.1.2 PM2.5**

**A.1.2.1 AIC-B**

Tabela A.5: Evolução dos indicadores de desempenho do modelo AIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 15 | bi+ii+c+gi+ej+rc+e+b+af+bg+fi+ab+aa+ce+aj+ci+g+i+cc | 0.00121        | -2585.6 | -2490.9 | 0.9517         |
| 19 | 16 | bi+ii+c+gi+ej+rc+e+b+af+bg+fi+ab+aa+ce+aj+ci+g+i    | 0.00121        | -2587.5 | -2497.3 | 0.9517         |
| 18 | 13 | bi+ii+c+gi+ej+rc+e+b+af+bg+fi+ab+aa+ce+aj+ci+g      | 0.00136        | -2506.6 | -2420.9 | 0.9453         |
| 17 | 14 | bi+ii+c+gi+ej+rc+e+b+af+bg+fi+ab+aa+ce+aj+ci        | 0.00137        | -2505.5 | -2424.3 | 0.9451         |
| 16 | 14 | bi+ii+c+gi+ej+rc+e+b+af+bg+fi+ab+aa+ce+aj           | 0.00137        | -2507.5 | -2430.8 | 0.9451         |
| 15 | 14 | bi+ii+c+gi+ej+rc+e+b+af+bg+fi+ab+aa+ce              | 0.00136        | -2509.1 | -2437.0 | 0.9451         |
| 14 | 12 | bi+ii+c+gi+ej+rc+e+b+af+bg+fi+ab+aa                 | 0.00151        | -2443.6 | -2376.0 | 0.9393         |
| 13 | 12 | bi+ii+c+gi+ej+rc+e+b+af+bg+fi+ab                    | 0.00151        | -2441.9 | -2378.8 | 0.9389         |
| 12 | 7  | bi+ii+c+gi+ej+rc+e+b+af+bg+fi                       | 0.00161        | -2400.1 | -2341.5 | 0.9348         |
| 11 | 7  | bi+ii+c+gi+ej+rc+e+b+af+bg                          | 0.00161        | -2402.1 | -2348.0 | 0.9348         |
| 10 | 8  | bi+ii+c+gi+ej+rc+e+b+af                             | 0.00161        | -2404.0 | -2354.4 | 0.9348         |
| 9  | 7  | bi+ii+c+gi+ej+rc+e+b                                | 0.00171        | -2362.9 | -2317.8 | 0.9305         |
| 8  | 6  | bi+ii+c+gi+ej+rc+e                                  | 0.00180        | -2328.8 | -2288.2 | 0.9267         |
| 7  | 5  | bi+ii+c+gi+ej+rc                                    | 0.00185        | -2313.9 | -2277.8 | 0.9248         |
| 6  | 5  | bi+ii+c+gi+ej                                       | 0.00190        | -2296.2 | -2264.7 | 0.9226         |
| 5  | 4  | bi+ii+c+gi  | 0.00409        | -1781.1 | -1754.1 | 0.8329         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

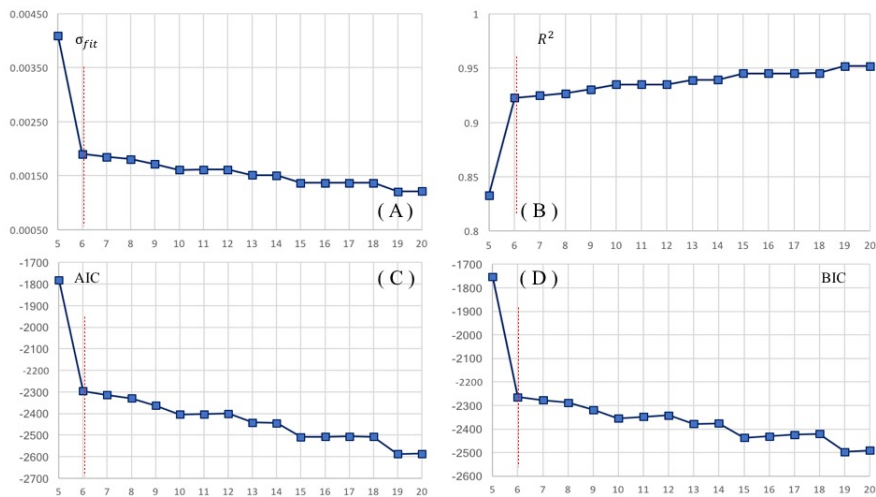


Figura A.9: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

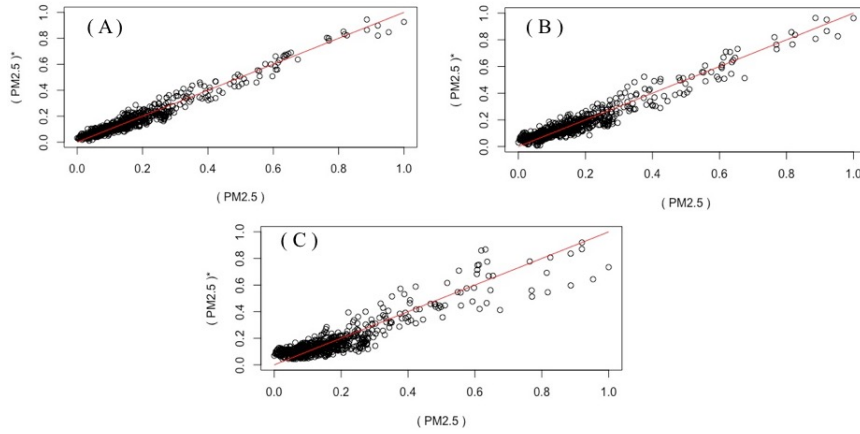


Figura A.10: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 6$ , (C) modelo deficitário com  $p = 5$  e modelo extremo com  $p = 5$ .

**A.1.2.2 AIC-BF**

Tabela A.6: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 19 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi+bg+aa+aj+i+ce+ci+g+ah | 0.00109        | -2653.1 | -2558.4 | 0.9563         |
| 19 | 16 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi+bg+aa+aj+i+ce+ci+g    | 0.00121        | -2587.5 | -2497.3 | 0.9517         |
| 18 | 16 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi+bg+aa+aj+i+ce+ci      | 0.00121        | -2588.7 | -2503.0 | 0.9516         |
| 17 | 13 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi+bg+aa+aj+i+ce         | 0.00133        | -2525.6 | -2444.4 | 0.9467         |
| 16 | 14 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi+bg+aa+aj+i            | 0.00147        | -2460.3 | -2383.6 | 0.9411         |
| 15 | 12 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi+bg+aa+aj              | 0.00150        | -2445.5 | -2373.3 | 0.9396         |
| 14 | 12 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi+bg+aa                 | 0.00151        | -2443.6 | -2376.0 | 0.9393         |
| 13 | 12 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi+bg                    | 0.00151        | -2441.9 | -2378.8 | 0.9389         |
| 12 | 11 | ii+bi+c+gi+ej+b+rc+e+af+ab+fi                       | 0.00152        | -2437.5 | -2378.9 | 0.9383         |
| 11 | 10 | ii+bi+c+gi+ej+b+rc+e+af+ab                          | 0.00158        | -2413.0 | -2358.9 | 0.9359         |
| 10 | 8  | ii+bi+c+gi+ej+b+rc+e+af                             | 0.00161        | -2404.0 | -2354.4 | 0.9348         |
| 9  | 7  | ii+bi+c+gi+ej+b+rc+e                                | 0.00171        | -2362.9 | -2317.8 | 0.9305         |
| 8  | 7  | ii+bi+c+gi+ej+b+rc                                  | 0.00173        | -2355.6 | -2315.0 | 0.9295         |
| 7  | 6  | ii+bi+c+gi+ej+b                                     | 0.00182        | -2322.7 | -2286.6 | 0.9258         |
| 6  | 5  | ii+bi+c+gi+ej                                       | 0.00190        | -2296.2 | -2264.7 | 0.9226         |
| 5  | 4  | ii+bi+c+gi  | 0.00409        | -1781.1 | -1754.1 | 0.8329         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

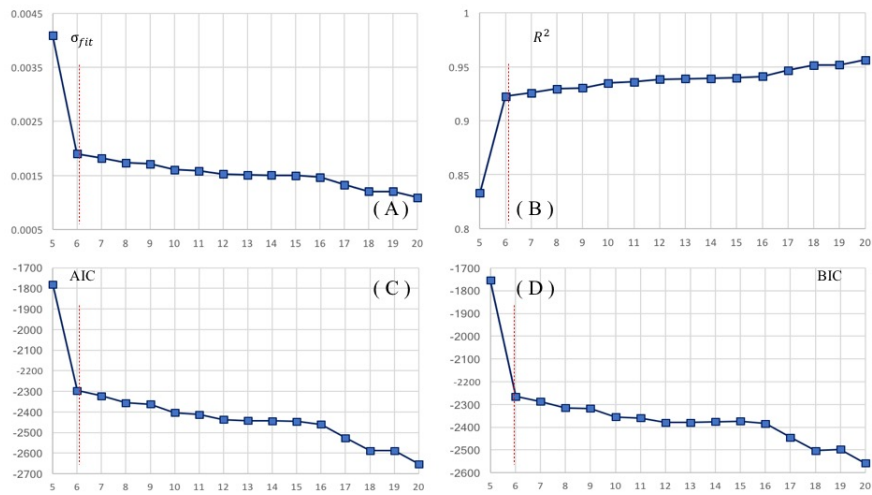


Figura A.11: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

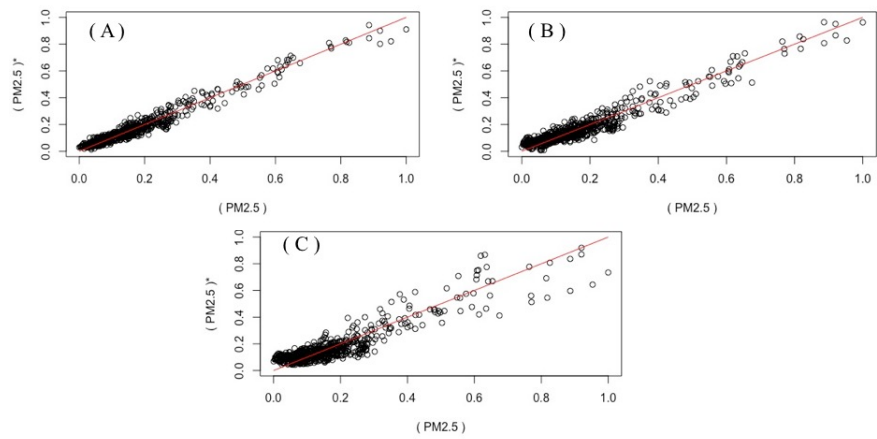


Figura A.12: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 6$ , (C) modelo deficitário com  $p = 5$  e modelo extremo com  $p = 5$ .



**A.1.2.3 BIC-B**

Tabela A.7: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 17 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa+bj+aj+cc+f | 0.00115        | -2618.9 | -2524.2 | 0.9540         |
| 19 | 14 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa+bj+aj+cc   | 0.00128        | -2548.1 | -2457.9 | 0.9488         |
| 18 | 15 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa+bj+aj      | 0.00128        | -2550.0 | -2464.3 | 0.9488         |
| 17 | 15 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa+bj         | 0.00128        | -2551.7 | -2470.5 | 0.9487         |
| 16 | 15 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa            | 0.00128        | -2553.1 | -2476.4 | 0.9487         |
| 15 | 11 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci               | 0.00133        | -2524.0 | -2451.9 | 0.9463         |
| 14 | 9  | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce                  | 0.00147        | -2460.1 | -2392.5 | 0.9407         |
| 13 | 9  | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg                     | 0.00157        | -2417.9 | -2354.8 | 0.9367         |
| 12 | 10 | ii+bi+c+b+rc+gi+af+ej+i+e+fi                        | 0.00157        | -2419.7 | -2361.1 | 0.9367         |
| 11 | 10 | ii+bi+c+b+rc+gi+af+ej+i+e                           | 0.00157        | -2420.9 | -2366.8 | 0.9366         |
| 10 | 7  | ii+bi+c+b+rc+gi+af+ej+i                             | 0.00161        | -2401.7 | -2352.1 | 0.9346         |
| 9  | 8  | ii+bi+c+b+rc+gi+af+ej                               | 0.00164        | -2391.2 | -2346.1 | 0.9334         |
| 8  | 6  | ii+bi+c+b+rc+gi+af                                  | 0.00271        | -2054.9 | -2014.4 | 0.8898         |
| 7  | 6  | ii+bi+c+b+rc+gi                                     | 0.00370        | -1846.7 | -1810.6 | 0.8493         |
| 6  | 5  | ii+bi+c+b+rc  | 0.00464        | -1696.0 | -1664.4 | 0.8109         |
| 5  | 4  | ii+bi+c+b   | 0.00501        | -1645.7 | -1618.7 | 0.7956         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

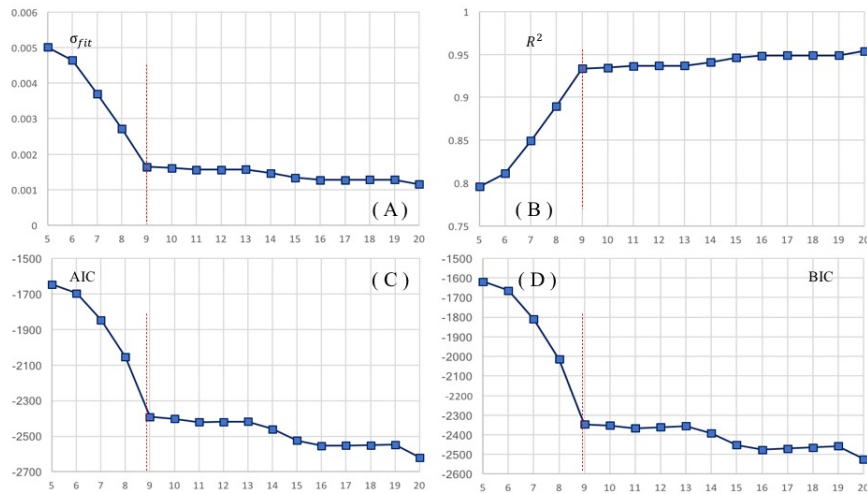


Figura A.13: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward.

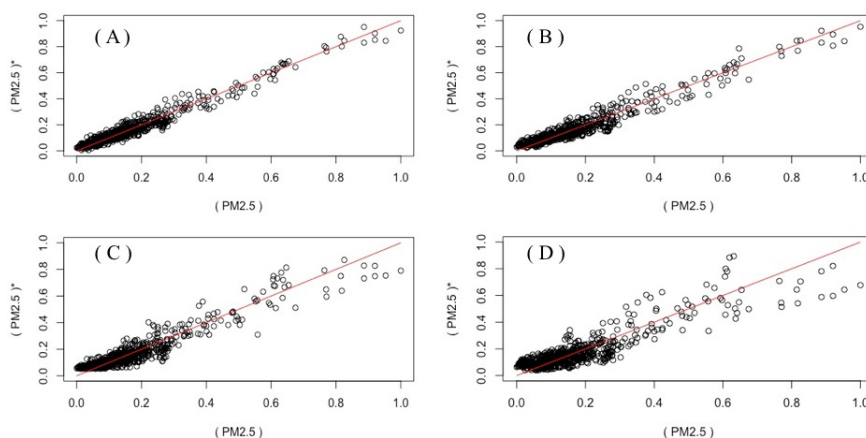


Figura A.14: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 9$ , (C) modelo deficitário com  $p = 8$  e (D) modelo extremo com  $p = 5$ .

**A.1.2.4 BIC-BF**

Tabela A.8: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 17 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa+bj+aj+cc+f | 0.00115        | -2618.9 | -2524.2 | 0.9540         |
| 19 | 14 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa+bj+aj+cc   | 0.00128        | -2548.1 | -2457.9 | 0.9488         |
| 18 | 15 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa+bj+aj      | 0.00128        | -2550.0 | -2464.3 | 0.9488         |
| 17 | 15 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa+bj         | 0.00128        | -2551.7 | -2470.5 | 0.9487         |
| 16 | 15 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci+aa            | 0.00128        | -2553.1 | -2476.4 | 0.9487         |
| 15 | 11 | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce+ci               | 0.00133        | -2524.0 | -2451.9 | 0.9463         |
| 14 | 9  | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg+ce                  | 0.00147        | -2460.1 | -2392.5 | 0.9407         |
| 13 | 9  | ii+bi+c+b+rc+gi+af+ej+i+e+fi+bg                     | 0.00157        | -2417.9 | -2354.8 | 0.9367         |
| 12 | 10 | ii+bi+c+b+rc+gi+af+ej+i+e+fi                        | 0.00157        | -2419.7 | -2361.1 | 0.9367         |
| 11 | 10 | ii+bi+c+b+rc+gi+af+ej+i+e                           | 0.00157        | -2420.9 | -2366.8 | 0.9366         |
| 10 | 7  | ii+bi+c+b+rc+gi+af+ej+i                             | 0.00161        | -2401.7 | -2352.1 | 0.9346         |
| 9  | 8  | ii+bi+c+b+rc+gi+af+ej                               | 0.00164        | -2391.2 | -2346.1 | 0.9334         |
| 8  | 6  | ii+bi+c+b+rc+gi+af                                  | 0.00271        | -2054.9 | -2014.4 | 0.8898         |
| 7  | 6  | ii+bi+c+b+rc+gi                                     | 0.00370        | -1846.7 | -1810.6 | 0.8493         |
| 6  | 5  | ii+bi+c+b+rc  | 0.00464        | -1696.0 | -1664.4 | 0.8109         |
| 5  | 4  | ii+bi+c+b   | 0.00501        | -1645.7 | -1618.7 | 0.7956         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

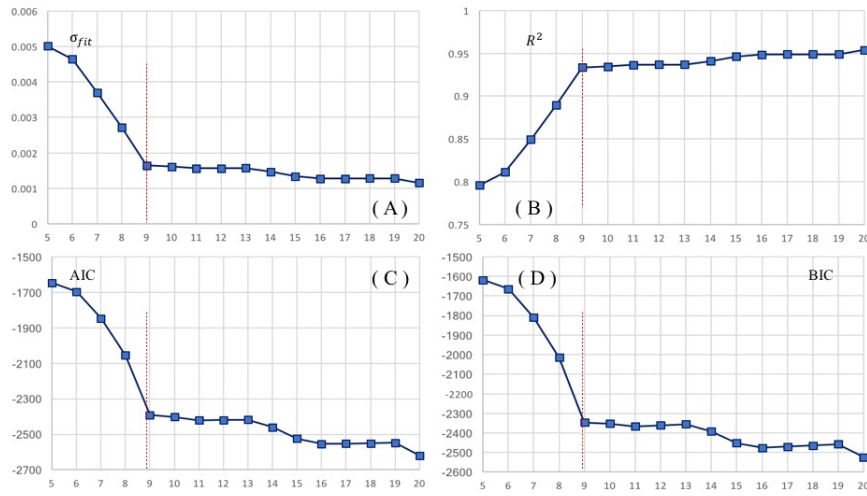


Figura A.15: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward-Forward.

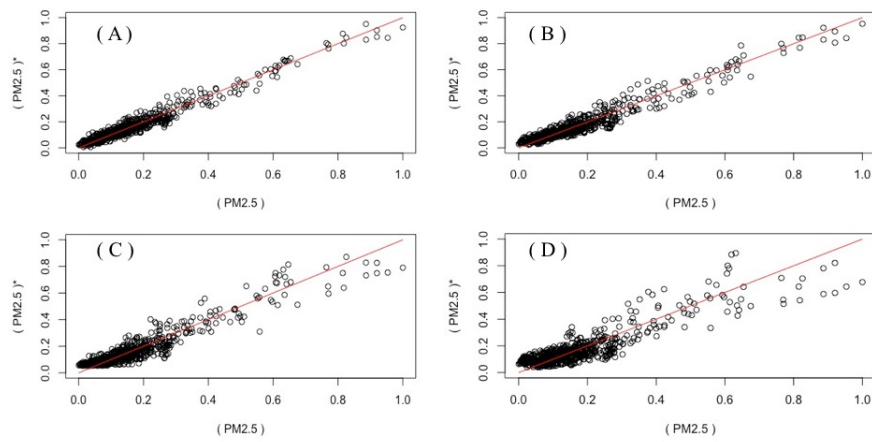


Figura A.16: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 9$ , (C) modelo deficitário com  $p = 8$  e (D) modelo extremo com  $p = 5$ .

**A.1.3 PM<sub>10</sub>**

**A.1.3.1 AIC-B**

Tabela A.9: Evolução dos indicadores de desempenho do modelo AIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 11 | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj+a+af+ag+cd+gi+h+aj+dg | 0.01011        | -1158.5 | -1063.7 | 0.7106         |
| 19 | 11 | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj+a+af+ag+cd+gi+h+aj    | 0.01010        | -1160.4 | -1070.2 | 0.7105         |
| 18 | 11 | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj+a+af+ag+cd+gi+h       | 0.01023        | -1152.7 | -1067.0 | 0.7063         |
| 17 | 12 | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj+a+af+ag+cd+gi         | 0.01144        | -1078.5 | -997.3  | 0.6711         |
| 16 | 13 | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj+a+af+ag+cd            | 0.01143        | -1080.1 | -1003.5 | 0.6709         |
| 15 | 9  | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj+a+af+ag               | 0.01483        | -906.1  | -834.0  | 0.5724         |
| 14 | 8  | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj+a+af                  | 0.01487        | -905.3  | -837.6  | 0.5705         |
| 13 | 8  | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj+a                     | 0.01486        | -906.5  | -843.4  | 0.5701         |
| 12 | 9  | dd+df+bi+c+dj+aa+rc+ii+g+ab+bj                       | 0.01485        | -908.2  | -849.5  | 0.5698         |
| 11 | 9  | dd+df+bi+c+dj+aa+rc+ii+g+ab                          | 0.01498        | -903.1  | -848.9  | 0.5653         |
| 10 | 8  | dd+df+bi+c+dj+aa+rc+ii+g                             | 0.01543        | -884.2  | -834.6  | 0.5516         |
| 9  | 8  | dd+df+bi+c+dj+aa+rc+ii                               | 0.01545        | -884.2  | -839.1  | 0.5502         |
| 8  | 7  | dd+df+bi+c+dj+aa+rc                                  | 0.01562        | -878.2  | -837.6  | 0.5448         |
| 7  | 6  | dd+df+bi+c+dj+aa                                     | 0.01632        | -849.5  | -813.5  | 0.5236         |
| 6  | 5  | dd+df+bi+c+dj  | 0.01685        | -829.2  | -797.6  | 0.5074         |
| 5  | 3  | dd+df+bi+c   | 0.01735        | -810.2  | -783.2  | 0.4919         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

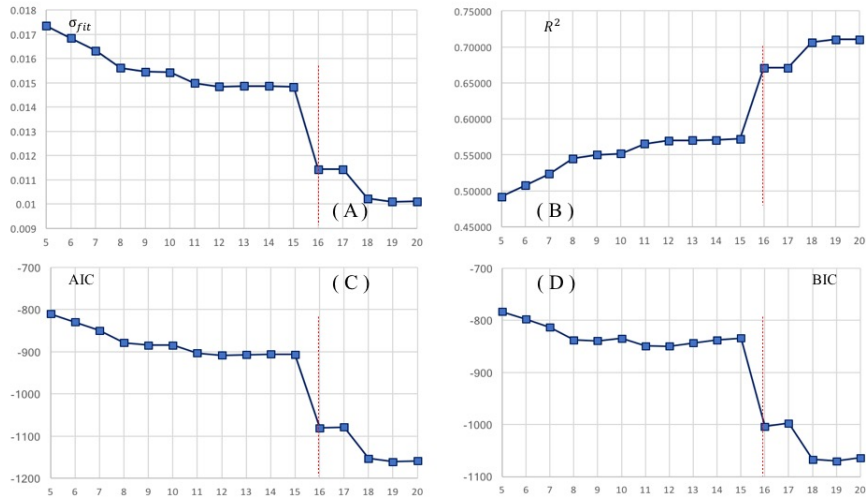


Figura A.17: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

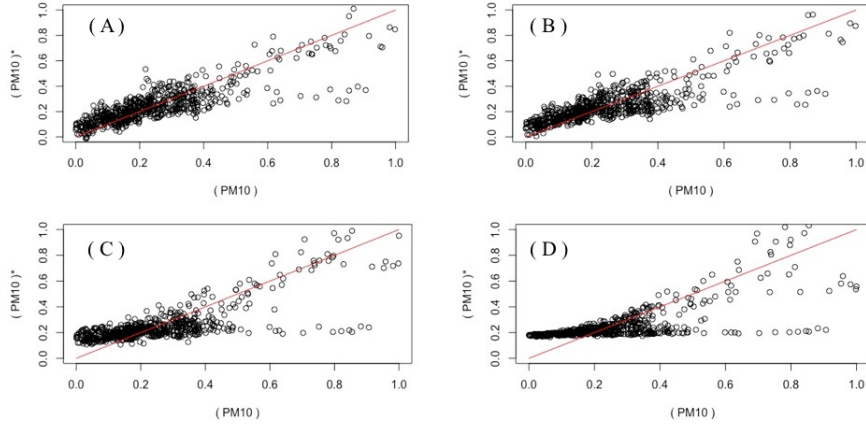


Figura A.18: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 16$ , (C) modelo deficitário com  $p = 15$  e (D) modelo extremo com  $p = 5$ .

**A.1.3.2 AIC-BF**

Tabela A.10: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 11 | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a+g+af+ag+h+cd+gi+aj+dg | 0.01011        | -1158.5 | -1063.7 | 0.7106         |
| 19 | 11 | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a+g+af+ag+h+cd+gi+aj    | 0.01010        | -1160.4 | -1070.2 | 0.7105         |
| 18 | 11 | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a+g+af+ag+h+cd+gi       | 0.01023        | -1152.7 | -1067.0 | 0.7063         |
| 17 | 14 | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a+g+af+ag+h+cd          | 0.01021        | -1154.7 | -1073.5 | 0.7063         |
| 16 | 9  | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a+g+af+ag+h             | 0.01468        | -912.0  | -835.3  | 0.5773         |
| 15 | 9  | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a+g+af+ag               | 0.01483        | -906.1  | -834.0  | 0.5724         |
| 14 | 8  | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a+g+af                  | 0.01487        | -905.3  | -837.6  | 0.5705         |
| 13 | 8  | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a+g                     | 0.01486        | -906.5  | -843.4  | 0.5701         |
| 12 | 9  | dd+df+bi+c+dj+aa+ii+ab+rc+bj+a                       | 0.01487        | -907.2  | -848.6  | 0.5692         |
| 11 | 9  | dd+df+bi+c+dj+aa+ii+ab+rc+bj                         | 0.01486        | -908.7  | -854.6  | 0.5689         |
| 10 | 9  | dd+df+bi+c+dj+aa+ii+ab+rc                            | 0.01497        | -904.4  | -854.8  | 0.5648         |
| 9  | 8  | dd+df+bi+c+dj+aa+ii+ab                               | 0.01560        | -878.0  | -832.9  | 0.5461         |
| 8  | 8  | dd+df+bi+c+dj+aa+ii                                  | 0.01614        | -855.9  | -815.3  | 0.5294         |
| 7  | 6  | dd+df+bi+c+dj+aa                                     | 0.01632        | -849.5  | -813.5  | 0.5236         |
| 6  | 5  | dd+df+bi+c+dj  | 0.01685        | -829.2  | -797.6  | 0.5074         |
| 5  | 3  | dd+df+bi+c   | 0.01735        | -810.2  | -783.2  | 0.4919         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

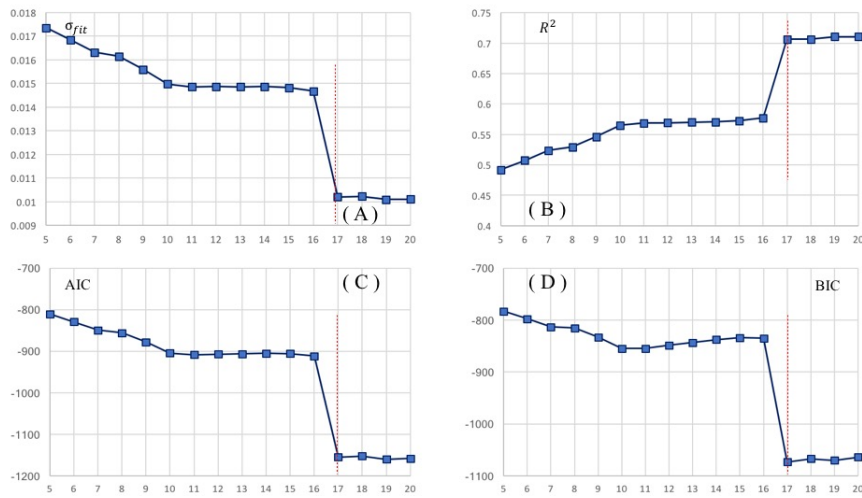


Figura A.19: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

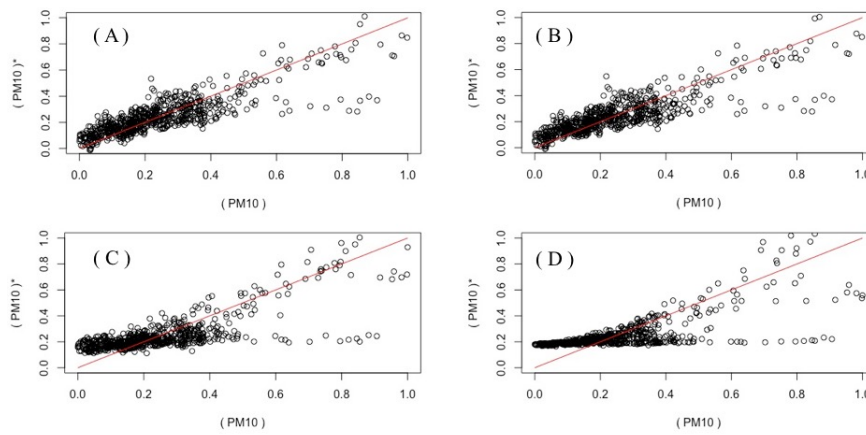


Figura A.20: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 17$ , (C) modelo deficitário com  $p = 16$  e (D) modelo extremo com  $p = 5$ .

**A.1.3.3 BIC-B**

Tabela A.11: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 13 | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa+fg+g+h+rh | 0.0101         | -1156.2 | -1061.4 | 0.7096         |
| 19 | 14 | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa+fg+g+h    | 0.0101         | -1158.0 | -1067.7 | 0.7095         |
| 18 | 13 | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa+fg+g      | 0.0112         | -1086.3 | -1000.6 | 0.6758         |
| 17 | 7  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa+fg        | 0.0115         | -1071.1 | -989.9  | 0.6674         |
| 16 | 7  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa           | 0.0115         | -1072.9 | -996.2  | 0.6673         |
| 15 | 11 | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj              | 0.0115         | -1071.7 | -999.5  | 0.6657         |
| 14 | 9  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd                 | 0.0119         | -1053.4 | -985.7  | 0.6555         |
| 13 | 8  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab                    | 0.0149         | -900.7  | -837.6  | 0.5664         |
| 12 | 5  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi                       | 0.566          | -883.0  | -824.4  | 0.5535         |
| 11 | 7  | dd+bi+df+c+rc+dj+af+bj+ag+ii                          | 0.0154         | -881.9  | -827.8  | 0.5514         |
| 10 | 6  | dd+bi+df+c+rc+dj+af+bj+ag                             | 0.0155         | -880.6  | -830.9  | 0.5491         |
| 9  | 7  | dd+bi+df+c+rc+dj+af+bj                                | 0.0154         | -882.6  | -837.5  | 0.5491         |
| 8  | 7  | dd+bi+df+c+rc+dj+af                                   | 0.0156         | -878.6  | -838.0  | 0.5451         |
| 7  | 5  | dd+bi+df+c+rc+dj                                      | 0.0159         | -863.5  | -827.4  | 0.5334         |
| 6  | 4  | dd+bi+df+c+rc   | 0.0161         | -857.6  | -826.0  | 0.5279         |
| 5  | 3  | dd+bi+df+c  | 0.0173         | -810.2  | -783.1  | 0.4919         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

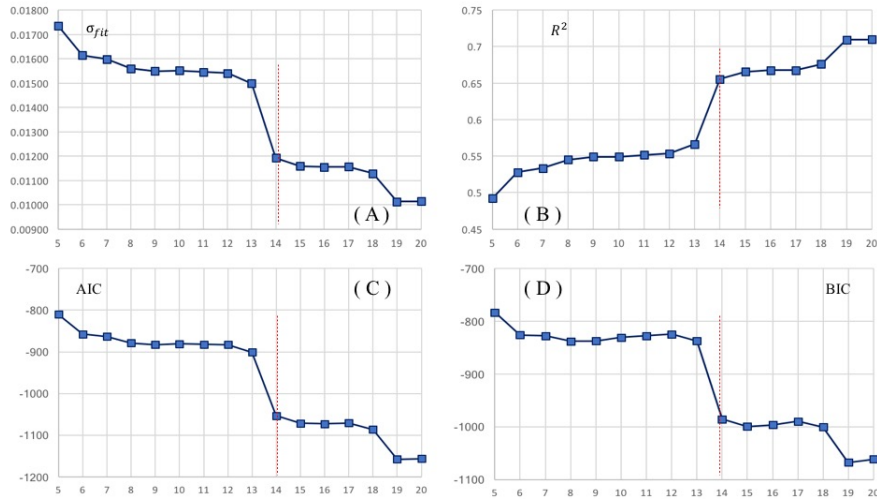


Figura A.21: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

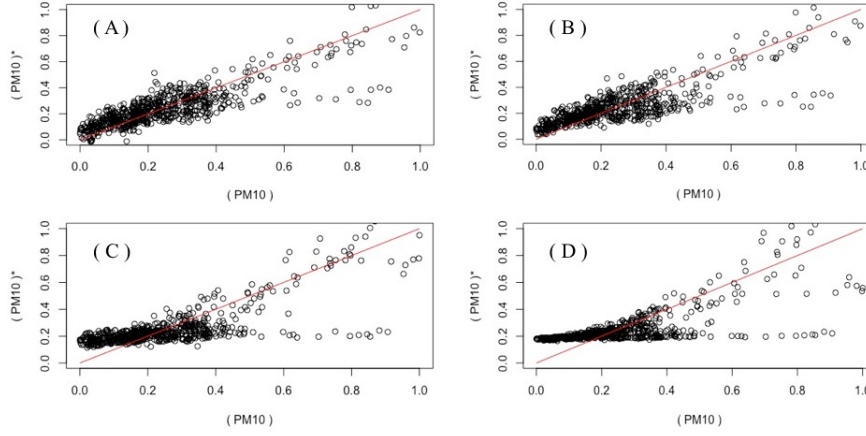


Figura A.22: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 14$ , (C) modelo deficitário com  $p = 13$  e (D) modelo extremo com  $p = 5$ .

**A.1.3.4 BIC-BF**

Tabela A.12: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 13 | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa+fg+g+h+rh | 0.0101         | -1156.2 | -1061.4 | 0.7096         |
| 19 | 14 | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa+fg+g+h    | 0.0101         | -1158.0 | -1067.7 | 0.7095         |
| 18 | 13 | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa+fg+g      | 0.0112         | -1086.3 | -1000.6 | 0.6758         |
| 17 | 7  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa+fg        | 0.0115         | -1071.1 | -989.9  | 0.6674         |
| 16 | 7  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj+aa           | 0.0115         | -1072.9 | -996.2  | 0.6673         |
| 15 | 11 | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd+aj              | 0.0115         | -1071.7 | -999.5  | 0.6657         |
| 14 | 9  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab+cd                 | 0.0119         | -1053.4 | -985.7  | 0.6555         |
| 13 | 8  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi+ab                    | 0.0149         | -900.7  | -837.6  | 0.5664         |
| 12 | 5  | dd+bi+df+c+rc+dj+af+bj+ag+ii+gi                       | 0.566          | -883.0  | -824.4  | 0.5535         |
| 11 | 7  | dd+bi+df+c+rc+dj+af+bj+ag+ii                          | 0.0154         | -881.9  | -827.8  | 0.5514         |
| 10 | 6  | dd+bi+df+c+rc+dj+af+bj+ag                             | 0.0155         | -880.6  | -830.9  | 0.5491         |
| 9  | 7  | dd+bi+df+c+rc+dj+af+bj                                | 0.0154         | -882.6  | -837.5  | 0.5491         |
| 8  | 7  | dd+bi+df+c+rc+dj+af                                   | 0.0156         | -878.6  | -838.0  | 0.5451         |
| 7  | 5  | dd+bi+df+c+rc+dj                                      | 0.0159         | -863.5  | -827.4  | 0.5334         |
| 6  | 4  | dd+bi+df+c+rc   | 0.0161         | -857.6  | -826.0  | 0.5279         |
| 5  | 3  | dd+bi+df+c  | 0.0173         | -810.2  | -783.1  | 0.4919         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;



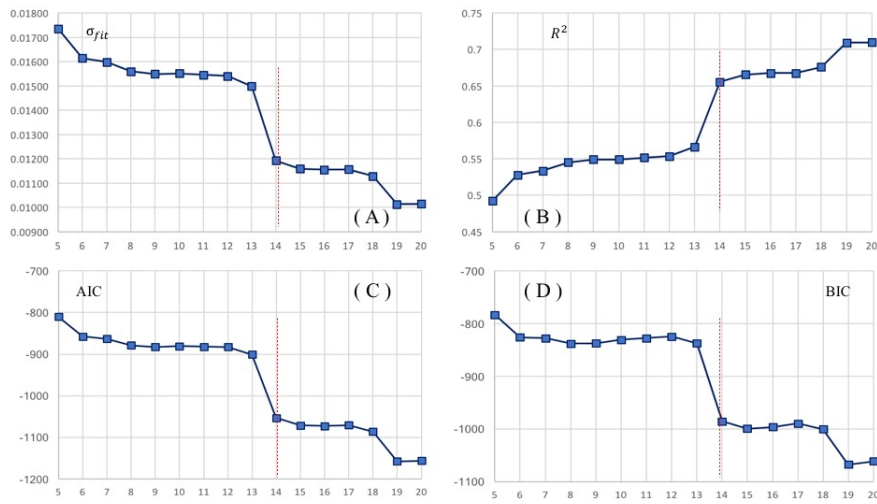


Figura A.23: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

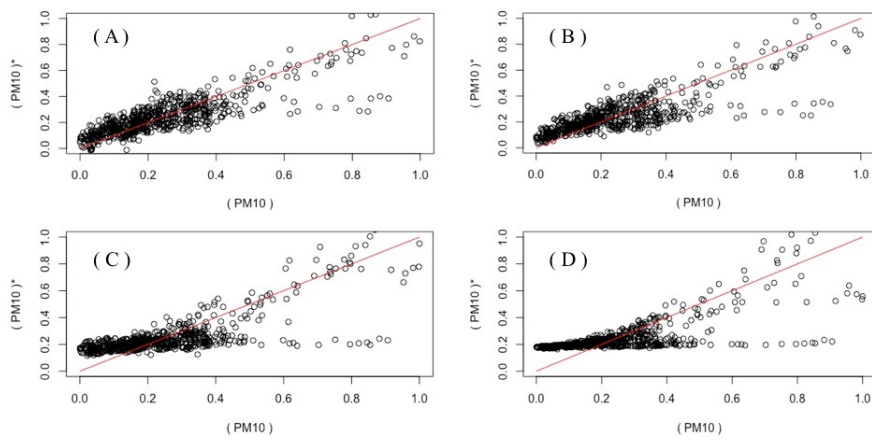


Figura A.24: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 14$ , (C) modelo deficitário com  $p = 13$  e (D) modelo extremo com  $p = 5$ .

### A.1.4 Dióxido de enxofre

#### A.1.4.1 AIC-B

Tabela A.13: Evolução dos indicadores de desempenho do modelo AIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 13 | de+ab+dd+b+ai+dj+a+ej+j+ii+bb+ra+aj+aa+rj+ag+ei+be+ce | 0.00672        | -1432.6 | -1337.8 | 0.8351         |
| 19 | 13 | de+ab+dd+b+ai+dj+a+ej+j+ii+bb+ra+aj+aa+rj+ag+ei+be    | 0.00675        | -1431.0 | -1340.8 | 0.8342         |
| 18 | 12 | de+ab+dd+b+ai+dj+a+ej+j+ii+bb+ra+aj+aa+rj+ag+ei       | 0.00681        | -1426.5 | -1340.8 | 0.8326         |
| 17 | 11 | de+ab+dd+b+ai+dj+a+ej+j+ii+bb+ra+aj+aa+rj+ag          | 0.00709        | -1399.8 | -1318.6 | 0.8253         |
| 16 | 10 | de+ab+dd+b+ai+dj+a+ej+j+ii+bb+ra+aj+aa+rj             | 0.00740        | -1372.2 | -1295.6 | 0.8174         |
| 15 | 10 | de+ab+dd+b+ai+dj+a+ej+j+ii+bb+ra+aj+aa                | 0.00750        | -1364.5 | -1292.3 | 0.8148         |
| 14 | 11 | de+ab+dd+b+ai+dj+a+ej+j+ii+bb+ra+aj                   | 0.00753        | -1362.3 | -1294.7 | 0.8136         |
| 13 | 8  | de+ab+dd+b+ai+dj+a+ej+j+ii+bb+ra                      | 0.00832        | -1296.2 | -1233.1 | 0.7937         |
| 12 | 9  | de+ab+dd+b+ai+dj+a+ej+j+ii+bb                         | 0.00834        | -1295.5 | -1236.9 | 0.7929         |
| 11 | 7  | de+ab+dd+b+ai+dj+a+ej+j+ii                            | 0.00890        | -1253.3 | -1199.2 | 0.7788         |
| 10 | 7  | de+ab+dd+b+ai+dj+a+ej+j                               | 0.00889        | -1255.1 | -1205.5 | 0.7788         |
| 9  | 7  | de+ab+dd+b+ai+dj+a+ej                                 | 0.00888        | -1256.6 | -1211.5 | 0.7786         |
| 8  | 7  | de+ab+dd+b+ai+dj+a                                    | 0.00888        | -1257.4 | -1216.9 | 0.7782         |
| 7  | 7  | de+ab+dd+b+ai+dj                                      | 0.00890        | -1256.9 | -1220.9 | 0.7774         |
| 6  | 5  | de+ab+dd+b+ai   | 0.01162        | -1078.7 | -1047.1 | 0.7089         |
| 5  | 4  | de+ab+dd+b  | 0.01381        | -963.9  | -936.8  | 0.6537         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

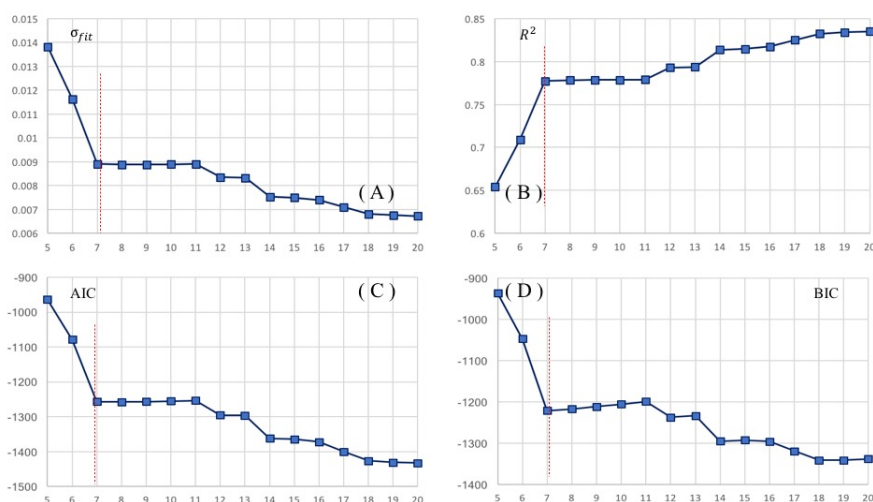


Figura A.25: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

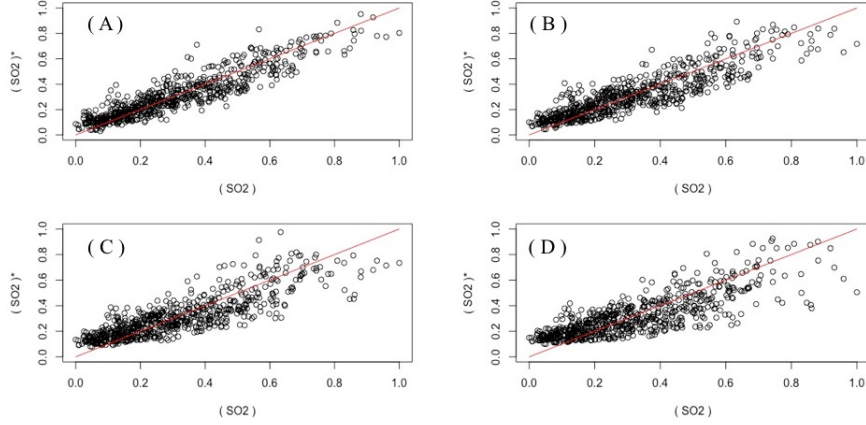


Figura A.26: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 7$ , (C) modelo deficitário com  $p = 6$  e (D) modelo extremo com  $p = 5$ .

**A.1.4.2 AIC-BF**

Tabela A.14: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 13 | de+ab+dd+b+ai+a+dj+j+ej+bb+ii+ag+ra+aa+aj+rj+ae+ei+be | 0.00675        | -1430.0 | -1335.3 | 0.8344         |
| 19 | 12 | de+ab+dd+b+ai+a+dj+j+ej+bb+ii+ag+ra+aa+aj+rj+ae+ei    | 0.00679        | -1427.4 | -1337.2 | 0.8333         |
| 18 | 12 | de+ab+dd+b+ai+a+dj+j+ej+bb+ii+ag+ra+aa+aj+rj+ae       | 0.00701        | -1406.4 | -1320.7 | 0.8275         |
| 17 | 11 | de+ab+dd+b+ai+a+dj+j+ej+bb+ii+ag+ra+aa+aj+rj          | 0.00709        | -1399.8 | -1318.6 | 0.8253         |
| 16 | 10 | de+ab+dd+b+ai+a+dj+j+ej+bb+ii+ag+ra+aa+aj             | 0.00720        | -1390.1 | -1313.5 | 0.8222         |
| 15 | 9  | de+ab+dd+b+ai+a+dj+j+ej+bb+ii+ag+ra+aa                | 0.00807        | -1314.8 | -1242.6 | 0.8005         |
| 14 | 11 | de+ab+dd+b+ai+a+dj+j+ej+bb+ii+ag+ra                   | 0.00808        | -1315.3 | -1247.6 | 0.8001         |
| 13 | 10 | de+ab+dd+b+ai+a+dj+j+ej+bb+ii+ag                      | 0.00811        | -1313.2 | -1250.1 | 0.7989         |
| 12 | 9  | de+ab+dd+b+ai+a+dj+j+ej+bb+ii                         | 0.00834        | -1295.5 | -1236.9 | 0.7929         |
| 11 | 7  | de+ab+dd+b+ai+a+dj+j+ej+bb                            | 0.00854        | -1281.0 | -1226.9 | 0.7877         |
| 10 | 7  | de+ab+dd+b+ai+a+dj+j+ej                               | 0.00889        | -1255.1 | -1205.5 | 0.7788         |
| 9  | 7  | de+ab+dd+b+ai+a+dj+j                                  | 0.00889        | -1256.0 | -1210.9 | 0.7784         |
| 8  | 7  | de+ab+dd+b+ai+a+dj                                    | 0.00888        | -1257.4 | -1216.9 | 0.7782         |
| 7  | 5  | de+ab+dd+b+ai+a                                       | 0.0115         | -1084.5 | -1048.4 | 0.7123         |
| 6  | 5  | de+ab+dd+b+ai   | 0.0116         | -1078.7 | -1047.1 | 0.7089         |
| 5  | 4  | de+ab+dd+b  | 0.0138         | -963.9  | -936.8  | 0.6537         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

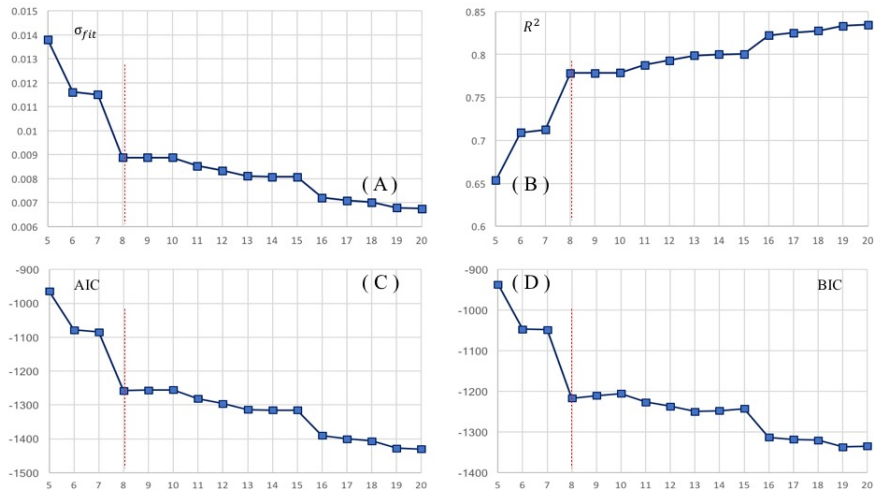


Figura A.27: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

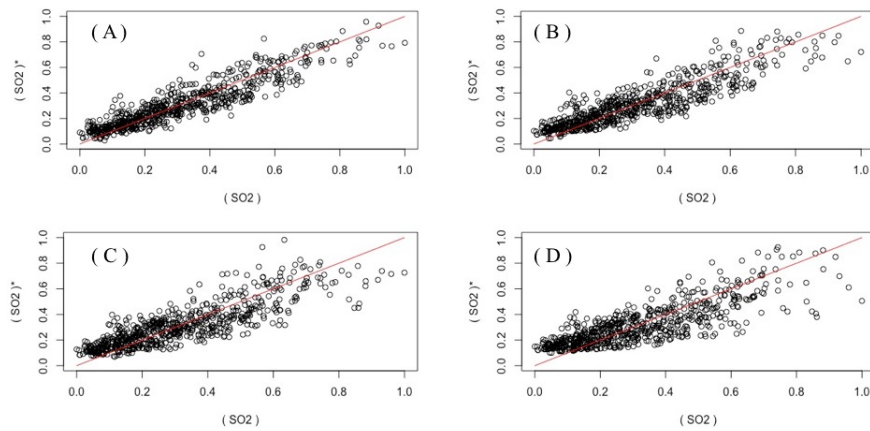


Figura A.28: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

A.1.4.3 BIC-B

Tabela A.15: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 20 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce+ah+dh+cd+hj | 0.00483        | -1654.5 | -1559.8 | 0.8814         |
| 19 | 19 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce+ah+dh+cd    | 0.00502        | -1629.8 | -1539.6 | 0.8767         |
| 18 | 18 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce+ah+dh       | 0.00633        | -1475.5 | -1389.8 | 0.8444         |
| 17 | 17 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce+ah          | 0.00641        | -1468.2 | -1387.0 | 0.8422         |
| 16 | 16 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce             | 0.00651        | -1458.2 | -1381.5 | 0.8393         |
| 15 | 14 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai                | 0.00670        | -1439.5 | -1367.3 | 0.8343         |
| 14 | 11 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg                   | 0.00739        | -1374.8 | -1307.1 | 0.8170         |
| 13 | 11 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh                      | 0.00740        | -1375.2 | -1312.1 | 0.8166         |
| 12 | 10 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag                         | 0.00906        | -1240.1 | -1181.4 | 0.7751         |
| 11 | 9  | ab+de+ae+dd+bb+b+j+ej+aj+rj                            | 0.00998        | -1176.2 | -1122.1 | 0.7519         |
| 10 | 8  | ab+de+ae+dd+bb+b+j+ej+aj                               | 0.0101         | -1162.9 | -1113.3 | 0.7462         |
| 9  | 7  | ab+de+ae+dd+bb+b+j+ej                                  | 0.0103         | -1154.5 | -1109.4 | 0.7423         |
| 8  | 8  | ab+de+ae+dd+bb+b+j                                     | 0.0103         | -1156.0 | -1115.4 | 0.7421         |
| 7  | 6  | ab+de+ae+dd+bb+b                                       | 0.0130         | -997.5  | -961.5  | 0.6725         |
| 6  | 3  | ab+de+ae+dd+bb   | 0.0212         | -674.5  | -643.0  | 0.4689         |
| 5  | 2  | ab+de+ae+dd  | 0.0214         | -669.3  | -642.3  | 0.4632         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

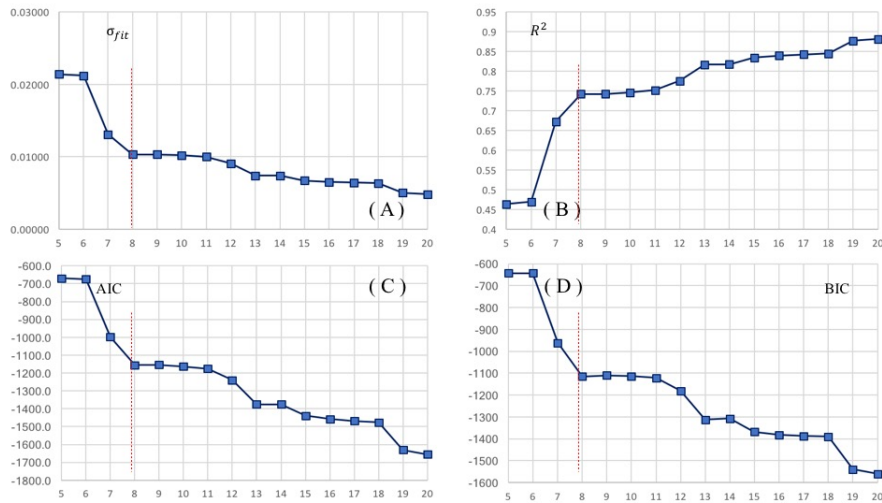


Figura A.29: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward.

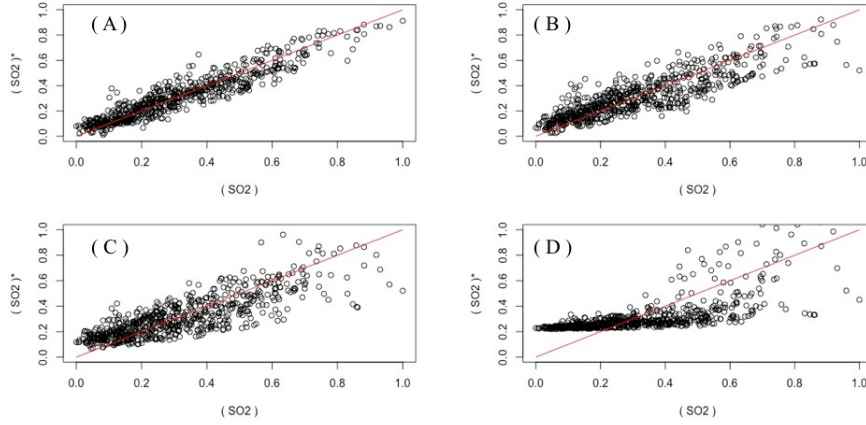


Figura A.30: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

**A.1.4.4 BIC-BF**

Tabela A.16: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 20 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce+ah+dh+cd+hj | 0.00483        | -1654.5 | -1559.8 | 0.8814         |
| 19 | 19 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce+ah+dh+cd    | 0.00502        | -1629.8 | -1539.6 | 0.8767         |
| 18 | 18 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce+ah+dh       | 0.00633        | -1475.5 | -1389.8 | 0.8444         |
| 17 | 17 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce+ah          | 0.00641        | -1468.2 | -1387.0 | 0.8422         |
| 16 | 16 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai+ce             | 0.00651        | -1458.2 | -1381.5 | 0.8393         |
| 15 | 14 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg+ai                | 0.00670        | -1439.5 | -1367.3 | 0.8343         |
| 14 | 11 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh+gg                   | 0.00739        | -1374.8 | -1307.1 | 0.8170         |
| 13 | 11 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag+gh                      | 0.00740        | -1375.2 | -1312.1 | 0.8166         |
| 12 | 10 | ab+de+ae+dd+bb+b+j+ej+aj+rj+ag                         | 0.00906        | -1240.1 | -1181.4 | 0.7751         |
| 11 | 9  | ab+de+ae+dd+bb+b+j+ej+aj+rj                            | 0.00998        | -1176.2 | -1122.1 | 0.7519         |
| 10 | 8  | ab+de+ae+dd+bb+b+j+ej+aj                               | 0.0101         | -1162.9 | -1113.3 | 0.7462         |
| 9  | 7  | ab+de+ae+dd+bb+b+j+ej                                  | 0.0103         | -1154.5 | -1109.4 | 0.7423         |
| 8  | 8  | ab+de+ae+dd+bb+b+j                                     | 0.0103         | -1156.0 | -1115.4 | 0.7421         |
| 7  | 6  | ab+de+ae+dd+bb+b                                       | 0.0130         | -997.5  | -961.5  | 0.6725         |
| 6  | 3  | ab+de+ae+dd+bb   | 0.0212         | -674.5  | -643.0  | 0.4689         |
| 5  | 2  | ab+de+ae+dd  | 0.0214         | -669.3  | -642.3  | 0.4632         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

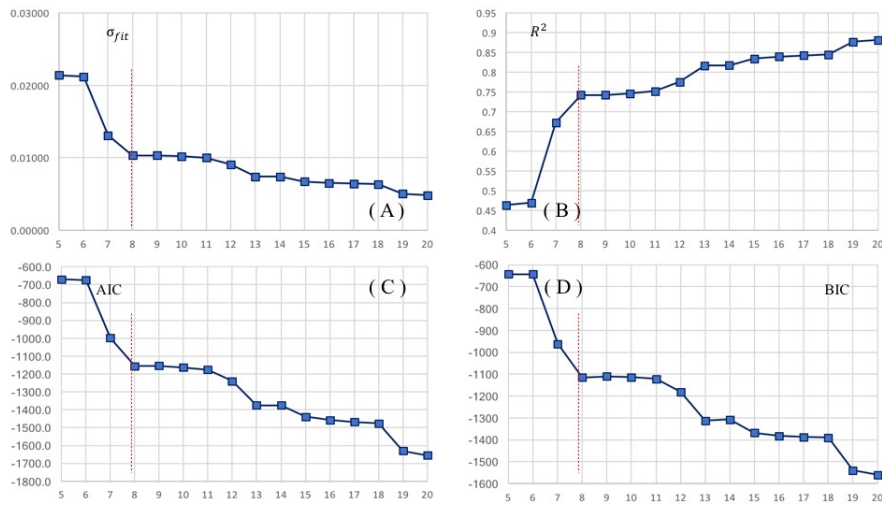


Figura A.31: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward-Forward.

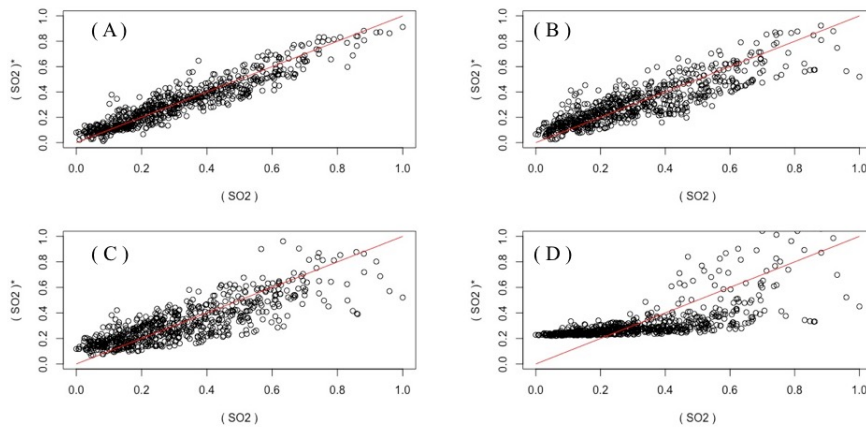


Figura A.32: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

### A.1.5 Monóxido de azoto

#### A.1.5.1 AIC-B

Tabela A.17: Evolução dos indicadores de desempenho do modelo AIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 18 | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh+ej+ae+be+i+ii+eh+ij+dh | 0.00527        | -1596.9 | -1502.2 | 0.5563         |
| 19 | 16 | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh+ej+ae+be+i+ii+eh+ij    | 0.00563        | -1552.8 | -1462.6 | 0.5248         |
| 18 | 13 | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh+ej+ae+be+i+ii+eh       | 0.00570        | -1546.0 | -1460.3 | 0.5186         |
| 17 | 11 | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh+ej+ae+be+i+ii          | 0.00626        | -1483.4 | -1402.3 | 0.4700         |
| 16 | 12 | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh+ej+ae+be+i             | 0.00625        | -1485.1 | -1408.5 | 0.4698         |
| 15 | 11 | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh+ej+ae+be               | 0.00639        | -1471.8 | -1399.7 | 0.4576         |
| 14 | 11 | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh+ej+ae                  | 0.00649        | -1462.8 | -1395.1 | 0.4486         |
| 13 | 9  | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh+ej                     | 0.00653        | -1459.0 | -1395.9 | 0.4438         |
| 12 | 9  | bd+bf+bb+dd+dj+ai+af+bi+b+h+rh                        | 0.00653        | -1459.9 | -1401.3 | 0.4429         |
| 11 | 8  | bd+bf+bb+dd+dj+ai+af+bi+b+h                           | 0.00661        | -1452.8 | -1398.7 | 0.4353         |
| 10 | 8  | bd+bf+bb+dd+dj+ai+af+bi+b                             | 0.00663        | -1452.0 | -1402.4 | 0.4330         |
| 9  | 9  | bd+bf+bb+dd+dj+ai+af+bi                               | 0.00682        | -1433.4 | -1388.3 | 0.4153         |
| 8  | 7  | bd+bf+bb+dd+dj+ai+af                                  | 0.00693        | -1424.2 | -1383.6 | 0.4054         |
| 7  | 6  | bd+bf+bb+dd+dj+ai                                     | 0.00770        | -1354.1 | -1318.0 | 0.3381         |
| 6  | 6  | bd+bf+bb+dd+dj  | 0.00794        | -1334.7 | -1303.1 | 0.3167         |
| 5  | 4  | bd+bf+bb+dd   | 0.00858        | -1283.6 | -1256.6 | 0.2606         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

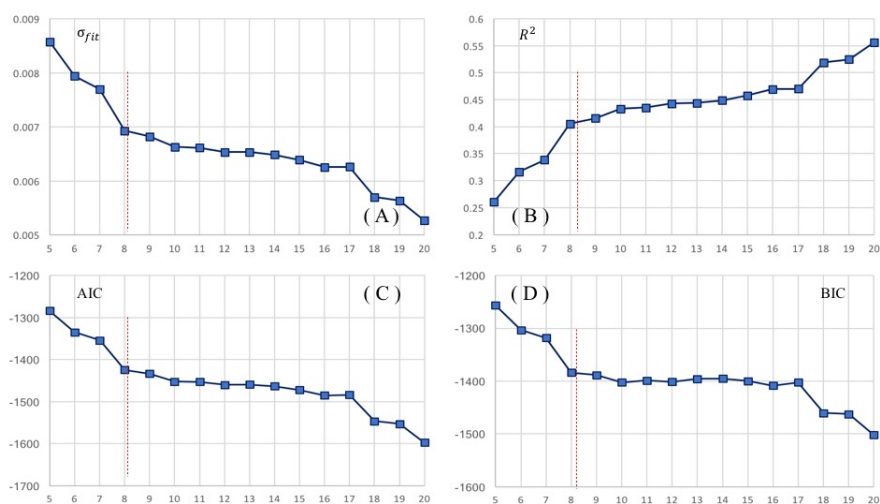


Figura A.33: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.



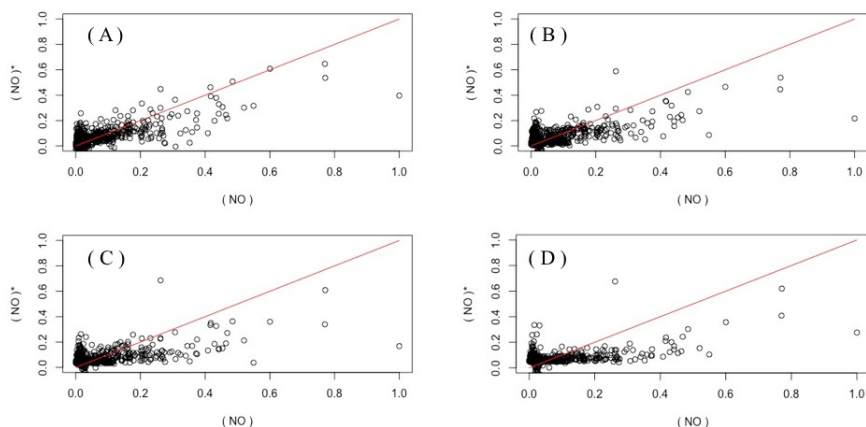


Figura A.34: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

### A.1.5.2 AIC-BF

Tabela A.18: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 18 | bd+bf+bb+dj+dd+ai+af+bi+b+h+i+ae+rh+ij+be+df+eh+ej+dh | 0.00509        | -1619.8 | -1525.1 | 0.5712         |
| 19 | 13 | bd+bf+bb+dj+dd+ai+af+bi+b+h+i+ae+rh+ij+be+df+eh+ej    | 0.00544        | -1576.5 | -1486.3 | 0.5413         |
| 18 | 15 | bd+bf+bb+dj+dd+ai+af+bi+b+h+i+ae+rh+ij+be+df+eh       | 0.00544        | -1577.6 | -1491.9 | 0.5407         |
| 17 | 12 | bd+bf+bb+dj+dd+ai+af+bi+b+h+i+ae+rh+ij+be+df          | 0.00559        | -1560.0 | -1478.8 | 0.5271         |
| 16 | 11 | bd+bf+bb+dj+dd+ai+af+bi+b+h+i+ae+rh+ij+be             | 0.00615        | -1496.0 | -1419.3 | 0.4782         |
| 15 | 13 | bd+bf+bb+dj+dd+ai+af+bi+b+h+i+ae+rh+ij                | 0.00617        | -1495.1 | -1422.9 | 0.4760         |
| 14 | 10 | bd+bf+bb+dj+dd+ai+af+bi+b+h+i+ae+rh                   | 0.00633        | -1479.2 | -1411.5 | 0.4618         |
| 13 | 8  | bd+bf+bb+dj+dd+ai+af+bi+b+h+i+ae                      | 0.00640        | -1472.6 | -1409.5 | 0.4550         |
| 12 | 8  | bd+bf+bb+dj+dd+ai+af+bi+b+h+i                         | 0.00639        | -1474.3 | -1415.7 | 0.4547         |
| 11 | 8  | bd+bf+bb+dj+dd+ai+af+bi+b+h                           | 0.00661        | -1452.8 | -1398.7 | 0.4353         |
| 10 | 8  | bd+bf+bb+dj+dd+ai+af+bi+b                             | 0.00663        | -1452.0 | -1402.4 | 0.4330         |
| 9  | 9  | bd+bf+bb+dj+dd+ai+af+bi                               | 0.00682        | -1433.4 | -1388.3 | 0.4153         |
| 8  | 7  | bd+bf+bb+dj+dd+ai+af                                  | 0.00693        | -1424.2 | -1383.6 | 0.4054         |
| 7  | 6  | bd+bf+bb+dj+dd+ai                                     | 0.00770        | -1354.1 | -1318.0 | 0.3381         |
| 6  | 6  | bd+bf+bb+dj+dd  | 0.00794        | -1334.7 | -1303.1 | 0.3167         |
| 5  | 4  | bd+bf+bb+dj   | 0.00841        | -1297.2 | -1270.1 | 0.2754         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

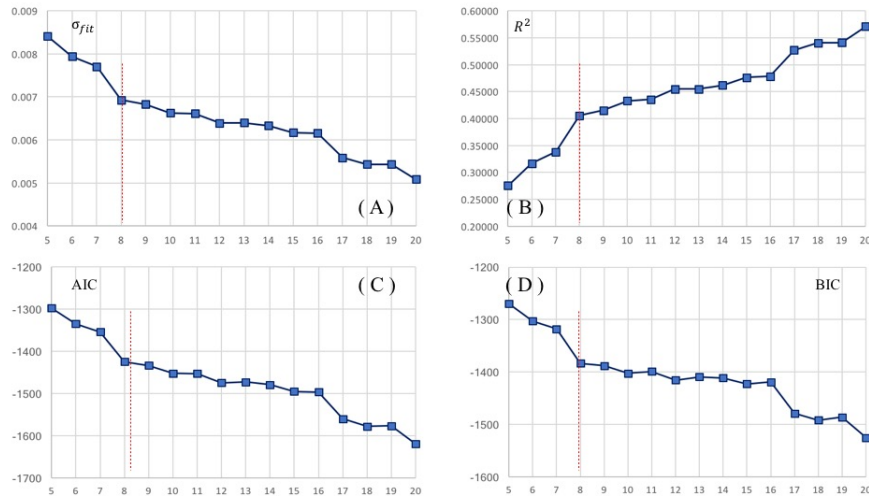


Figura A.35: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

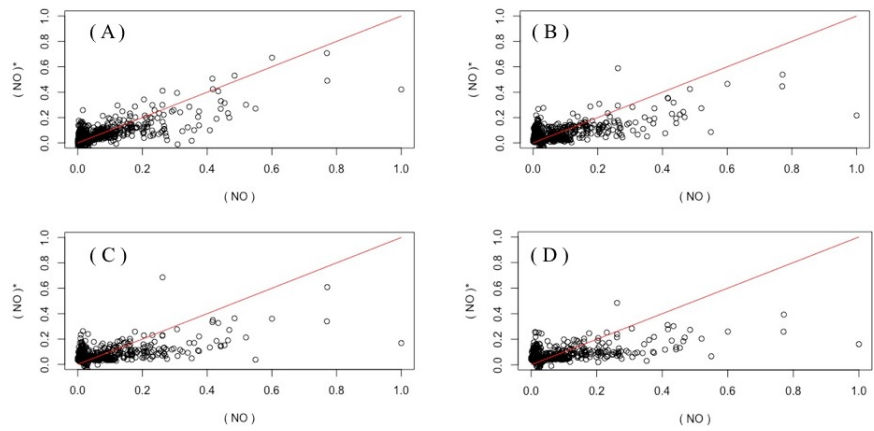


Figura A.36: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 8$ , (C) modelo deficitário com  $p = 7$  e (D) modelo extremo com  $p = 5$ .

**A.1.5.3 BIC-B**

Tabela A.19: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 17 | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh+bi+dh+ij+ej+b+eh+cj+f | 0.00544        | -1575.0 | -1480.3 | 0.5416         |
| 19 | 17 | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh+bi+dh+ij+ej+b+eh+cj   | 0.00544        | -1576.2 | -1485.9 | 0.5410         |
| 18 | 17 | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh+bi+dh+ij+ej+b+eh      | 0.00543        | -1577.9 | -1492.2 | 0.5409         |
| 17 | 16 | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh+bi+dh+ij+ej+b         | 0.00549        | -1572.3 | -1491.1 | 0.5357         |
| 16 | 15 | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh+bi+dh+ij+ej           | 0.00564        | -1554.2 | -1477.5 | 0.5216         |
| 15 | 14 | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh+bi+dh+ij              | 0.00569        | -1549.2 | -1477.0 | 0.5166         |
| 14 | 11 | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh+bi+dh                 | 0.00614        | -1499.6 | -1432.0 | 0.4780         |
| 13 | 7  | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh+bi                    | 0.00681        | -1431.4 | -1368.2 | 0.4205         |
| 12 | 8  | bd+dd+bf+dj+bb+af+ai+be+ae+h+rh                       | 0.00690        | -1423.0 | -1364.4 | 0.4115         |
| 11 | 8  | bd+dd+bf+dj+bb+af+ai+be+ae+h                          | 0.00690        | -1424.5 | -1370.3 | 0.4110         |
| 10 | 8  | bd+dd+bf+dj+bb+af+ai+be+ae                            | 0.00690        | -1425.3 | -1375.7 | 0.4100         |
| 9  | 7  | bd+dd+bf+dj+bb+af+ai+be                               | 0.00694        | -1422.2 | -1377.1 | 0.4055         |
| 8  | 7  | bd+dd+bf+dj+bb+af+ai                                  | 0.00693        | -1424.2 | -1383.6 | 0.4054         |
| 7  | 6  | bd+dd+bf+dj+bb+af                                     | 0.00703        | -1415.2 | -1379.1 | 0.3956         |
| 6  | 6  | bd+dd+bf+dj+bb  | 0.00794        | -1334.7 | -1303.1 | 0.3167         |
| 5  | 5  | bd+dd+bf+dj   | 0.00798        | -1332.1 | -1305.1 | 0.3121         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

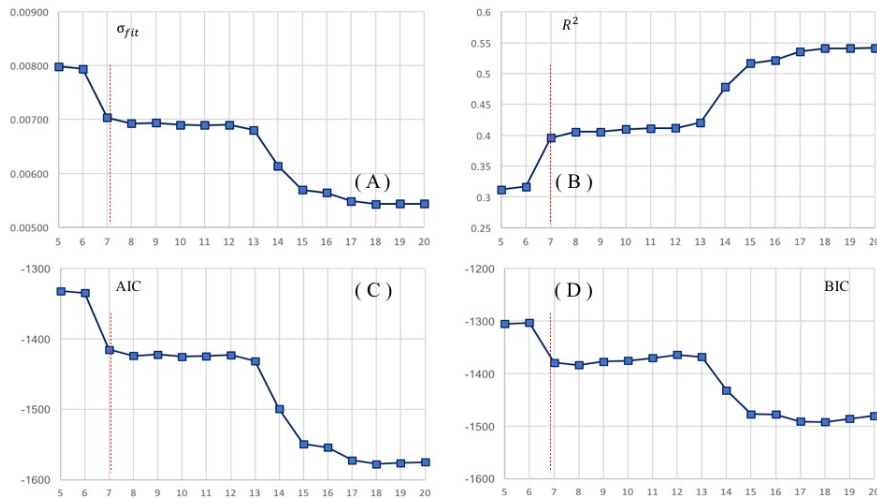


Figura A.37: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward.

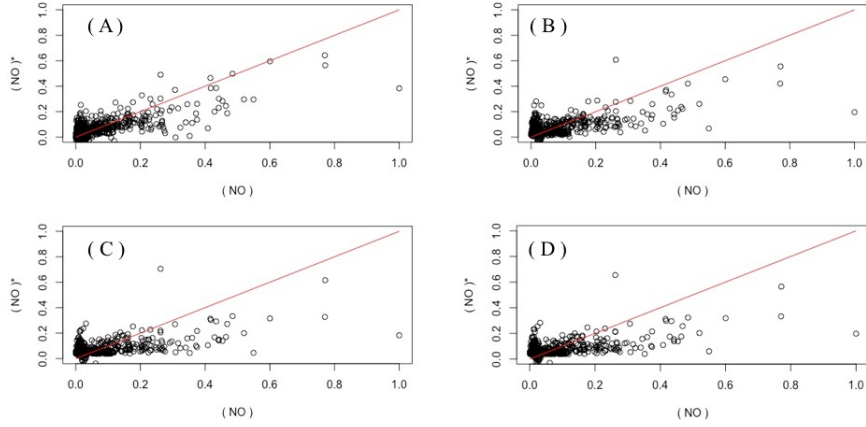


Figura A.38: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 7$ , (C) modelo deficitário com  $p = 6$  e (D) modelo extremo com  $p = 5$ .

**A.1.5.4 BIC-BF**

Tabela A.20: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 18 | bd+bf+bb+af+dj+dd+ai+ae+be+de+h+bi+dh+rh+ij+ej+eh+b+fj | 0.00521        | -1603.6 | -1508.8 | 0.5607         |
| 19 | 18 | bd+bf+bb+af+dj+dd+ai+ae+be+de+h+bi+dh+rh+ij+ej+eh+b    | 0.00529        | -1595.2 | -1505.0 | 0.5539         |
| 18 | 17 | bd+bf+bb+af+dj+dd+ai+ae+be+de+h+bi+dh+rh+ij+ej+eh      | 0.00541        | -1580.6 | -1494.9 | 0.5427         |
| 17 | 14 | bd+bf+bb+af+dj+dd+ai+ae+be+de+h+bi+dh+rh+ij+ej         | 0.00548        | -1572.6 | -1491.4 | 0.5358         |
| 16 | 14 | bd+bf+bb+af+dj+dd+ai+ae+be+de+h+bi+dh+rh+ij            | 0.00549        | -1572.4 | -1495.8 | 0.5344         |
| 15 | 12 | bd+bf+bb+af+dj+dd+ai+ae+be+de+h+bi+dh+rh               | 0.00576        | -1541.7 | -1469.5 | 0.5111         |
| 14 | 12 | bd+bf+bb+af+dj+dd+ai+ae+be+de+h+bi+dh                  | 0.00592        | -1523.9 | -1456.3 | 0.4965         |
| 13 | 10 | bd+bf+bb+af+dj+dd+ai+ae+be+de+h+bi                     | 0.00617        | -1497.4 | -1434.3 | 0.4747         |
| 12 | 9  | bd+bf+bb+af+dj+dd+ai+ae+be+de+h                        | 0.00619        | -1495.6 | -1436.9 | 0.4717         |
| 11 | 9  | bd+bf+bb+af+dj+dd+ai+ae+be+de                          | 0.00619        | -1497.3 | -1443.2 | 0.4715         |
| 10 | 8  | bd+bf+bb+af+dj+dd+ai+ae+be                             | 0.00690        | -1425.3 | -1375.7 | 0.4100         |
| 9  | 7  | bd+bf+bb+af+dj+dd+ai+ae                                | 0.00692        | -1423.7 | -1378.6 | 0.4068         |
| 8  | 7  | bd+bf+bb+af+dj+dd+ai                                   | 0.00693        | -1424.2 | -1383.6 | 0.4054         |
| 7  | 6  | bd+bf+bb+af+dj+dd                                      | 0.00703        | -1415.2 | -1379.1 | 0.3956         |
| 6  | 4  | bd+bf+bb+af+dj   | 0.00721        | -1399.3 | -1367.7 | 0.3793         |
| 5  | 5  | bd+bf+bb+af  | 0.00742        | -1381.4 | -1354.3 | 0.3607         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

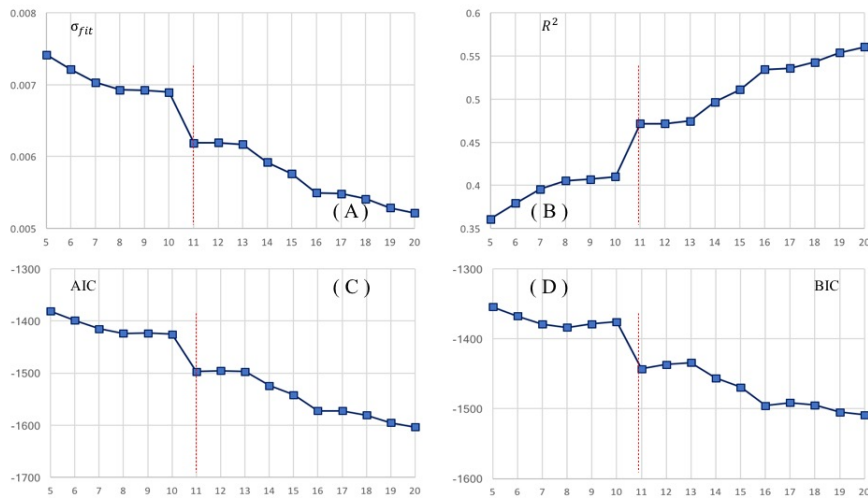


Figura A.39: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward-Forward.

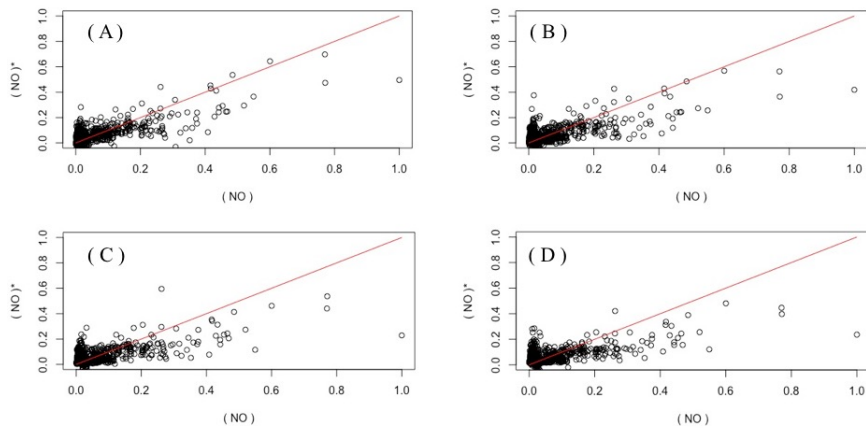


Figura A.40: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 11$ , (C) modelo deficitário com  $p = 10$  e (D) modelo extremo com  $p = 5$ .

**A.1.6 Amoníaco**

**A.1.6.1 AIC-B**

Tabela A.21: Evolução dos indicadores de desempenho do modelo AIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC    | BIC    | R <sup>2</sup> |
|----|----|--|----------------|--------|--------|----------------|
| 20 | 20 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg+aj+ee+ce+cd | 0.0162         | -841.1 | -746.3 | 0.6931         |
| 19 | 15 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg+aj+ee+ce    | 0.0194         | -720.3 | -630.1 | 0.6316         |
| 18 | 15 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg+aj+ee       | 0.0197         | -710.9 | -625.2 | 0.6253         |
| 17 | 15 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg+aj          | 0.0198         | -706.7 | -625.5 | 0.6218         |
| 16 | 15 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg             | 0.0208         | -674.7 | -598.1 | 0.6022         |
| 15 | 14 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi                | 0.0211         | -666.7 | -594.5 | 0.5962         |
| 14 | 13 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff                   | 0.0213         | -661.7 | -594.0 | 0.5919         |
| 13 | 12 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de                      | 0.0229         | -615.9 | -552.7 | 0.5619         |
| 12 | 11 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f                         | 0.0255         | -544.1 | -485.4 | 0.5110         |
| 11 | 11 | ad+aa+a+bi+ab+dd+b+ai+bb+d                           | 0.0272         | -500.0 | -445.8 | 0.4763         |
| 10 | 9  | ad+aa+a+bi+ab+dd+b+ai+bb                             | 0.0278         | -487.2 | -437.6 | 0.4647         |
| 9  | 9  | ad+aa+a+bi+ab+dd+b+ai                                | 0.0286         | -469.2 | -424.1 | 0.4485         |
| 8  | 8  | ad+aa+a+bi+ab+dd+b                                   | 0.0294         | -452.4 | -411.8 | 0.4328         |
| 7  | 6  | ad+aa+a+bi+ab+dd                                     | 0.0308         | -422.0 | -385.9 | 0.4048         |
| 6  | 5  | ad+aa+a+bi+ab  | 0.0388         | -267.8 | -236.2 | 0.2491         |
| 5  | 5  | ad+aa+a+bi   | 0.0390         | -265.0 | -237.9 | 0.2437         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

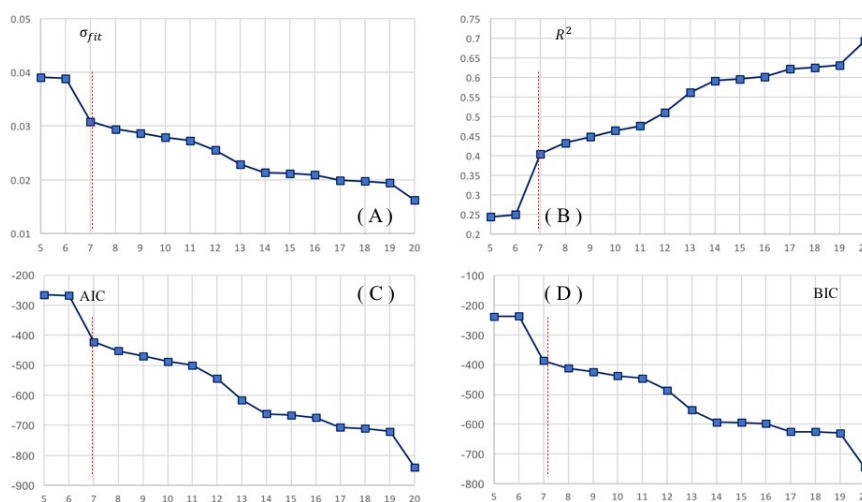


Figura A.41: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

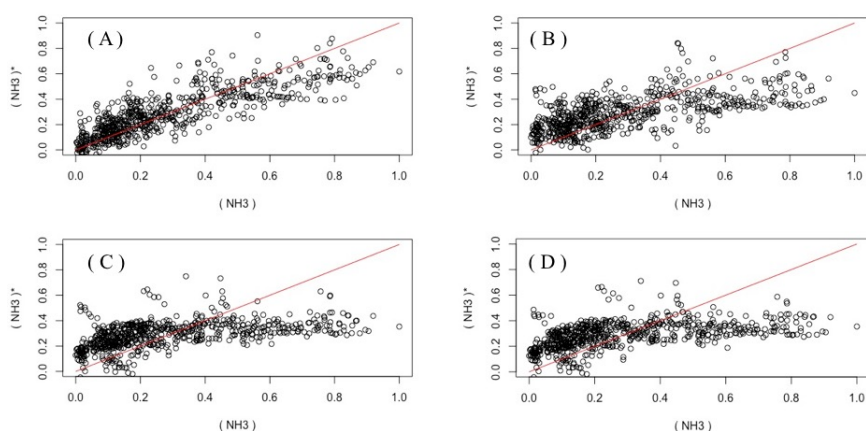


Figura A.42: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 17$ , (C) modelo deficitário com  $p = 16$  e (D) modelo extremo com  $p = 5$ .

### A.1.6.2 AIC-BF

Tabela A.22: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC    | BIC    | R <sup>2</sup> |
|----|----|--|----------------|--------|--------|----------------|
| 20 | 20 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg+aj+ee+ce+cd | 0.0162         | -841.1 | -746.3 | 0.6931         |
| 19 | 15 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg+aj+ee+ce    | 0.0194         | -720.3 | -630.1 | 0.6316         |
| 18 | 15 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg+aj+ee       | 0.0197         | -710.9 | -625.2 | 0.6253         |
| 17 | 15 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg+aj          | 0.0198         | -706.7 | -625.5 | 0.6218         |
| 16 | 15 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi+fg             | 0.0208         | -674.7 | -598.1 | 0.6022         |
| 15 | 14 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff+fi                | 0.0211         | -666.7 | -594.5 | 0.5962         |
| 14 | 13 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de+ff                   | 0.0213         | -661.7 | -594.0 | 0.5919         |
| 13 | 12 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f+de                      | 0.0229         | -615.9 | -552.7 | 0.5619         |
| 12 | 11 | ad+aa+a+bi+ab+dd+b+ai+bb+d+f                         | 0.0255         | -544.1 | -485.4 | 0.5110         |
| 11 | 11 | ad+aa+a+bi+ab+dd+b+ai+bb+d                           | 0.0272         | -500.0 | -445.8 | 0.4763         |
| 10 | 9  | ad+aa+a+bi+ab+dd+b+ai+bb                             | 0.0278         | -487.2 | -437.6 | 0.4647         |
| 9  | 9  | ad+aa+a+bi+ab+dd+b+ai                                | 0.0286         | -469.2 | -424.1 | 0.4485         |
| 8  | 8  | ad+aa+a+bi+ab+dd+b                                   | 0.0294         | -452.4 | -411.8 | 0.4328         |
| 7  | 6  | ad+aa+a+bi+ab+dd                                     | 0.0308         | -422.0 | -385.9 | 0.4048         |
| 6  | 5  | ad+aa+a+bi+ab  | 0.0388         | -267.8 | -236.2 | 0.2491         |
| 5  | 5  | ad+aa+a+bi   | 0.0390         | -265.0 | -237.9 | 0.2437         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

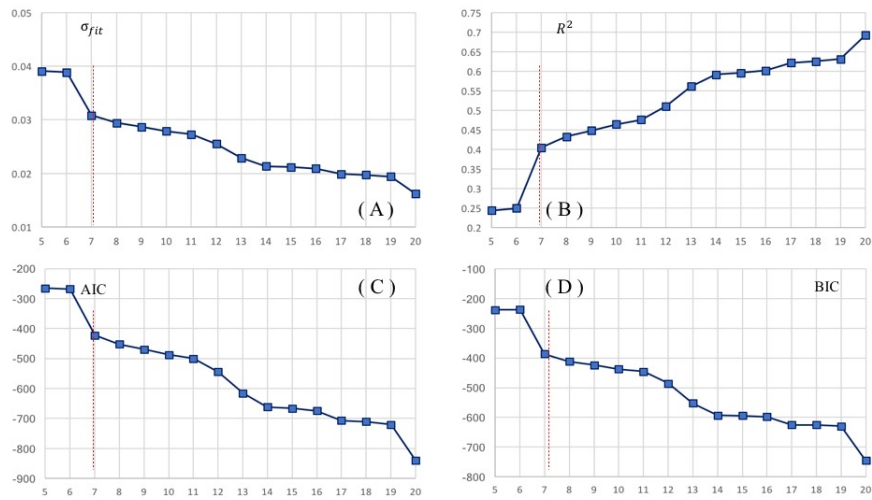


Figura A.43: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

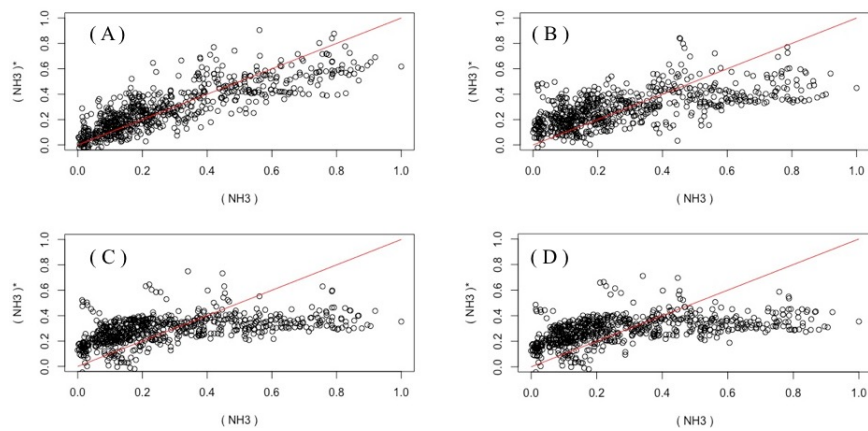


Figura A.44: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 17$ , (C) modelo deficitário com  $p = 16$  e (D) modelo extremo com  $p = 5$ .



A.1.6.3 BIC-B

Tabela A.23: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC    | BIC    | R <sup>2</sup> |
|----|----|--|----------------|--------|--------|----------------|
| 20 | 18 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee+ff+cd+j+gj | 0.0158         | -857.1 | -762.4 | 0.7003         |
| 19 | 17 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee+ff+cd+j    | 0.0169         | -813.3 | -723.1 | 0.6792         |
| 18 | 17 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee+ff+cd      | 0.0170         | -809.4 | -723.7 | 0.6764         |
| 17 | 14 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee+ff         | 0.0199         | -703.5 | -622.3 | 0.6200         |
| 16 | 15 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee            | 0.0216         | -650.6 | -573.9 | 0.5876         |
| 15 | 14 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce               | 0.0217         | -647.4 | -575.2 | 0.5844         |
| 14 | 12 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj                  | 0.0226         | -623.2 | -555.5 | 0.5679         |
| 13 | 12 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd                     | 0.0233         | -604.0 | -540.8 | 0.5541         |
| 12 | 11 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f                        | 0.0234         | -600.7 | -542.0 | 0.5505         |
| 11 | 11 | ad+aa+de+bb+a+bi+ai+b+fg+dd                          | 0.0246         | -569.1 | -515.0 | 0.5275         |
| 10 | 10 | ad+aa+de+bb+a+bi+ai+b+fg                             | 0.0300         | -435.6 | -386.0 | 0.4219         |
| 9  | 9  | ad+aa+de+bb+a+bi+ai+b                                | 0.0308         | -420.5 | -375.4 | 0.4070         |
| 8  | 8  | ad+aa+de+bb+a+bi+ai                                  | 0.0311         | -414.0 | -373.4 | 0.3995         |
| 7  | 6  | ad+aa+de+bb+a+bi                                     | 0.0341         | -353.7 | -317.6 | 0.3412         |
| 6  | 6  | ad+aa+de+bb+a  | 0.0340         | -355.4 | -323.9 | 0.3409         |
| 5  | 4  | ad+aa+de+bb  | 0.0387         | -269.8 | -242.7 | 0.2491         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

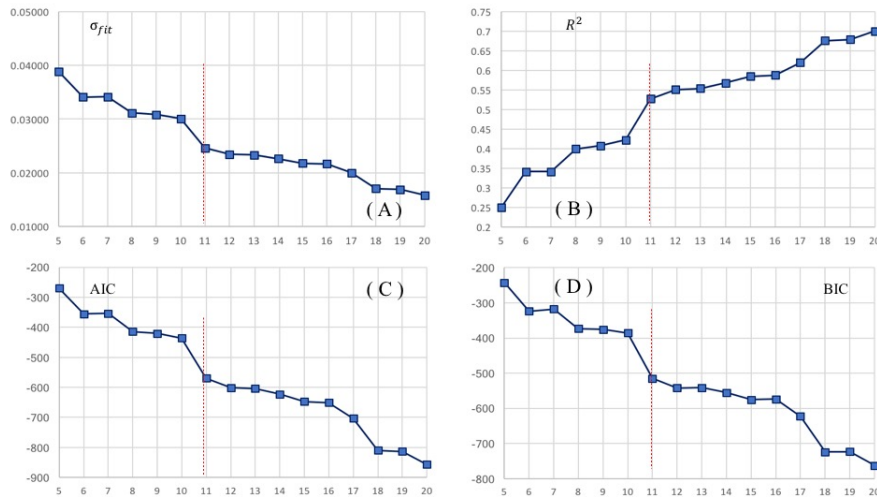


Figura A.45: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward.

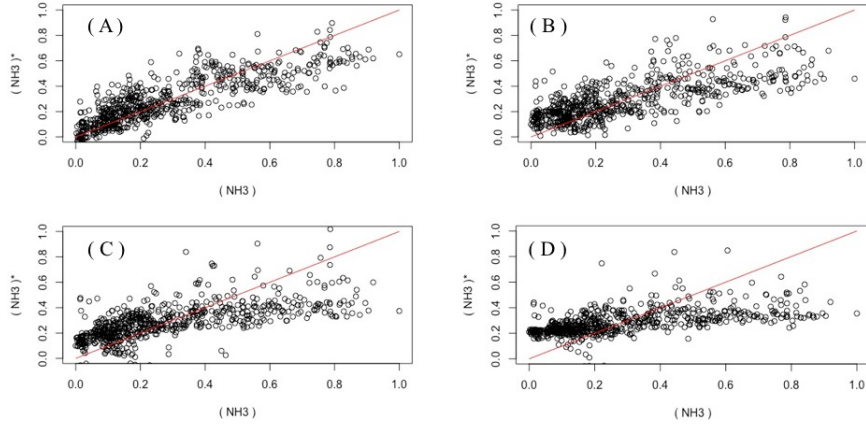


Figura A.46: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 15$ , (C) modelo deficitário com  $p = 14$  e (D) modelo extremo com  $p = 5$ .

**A.1.6.4 BIC-BF**

Tabela A.24: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC    | BIC    | R <sup>2</sup> |
|----|----|--|----------------|--------|--------|----------------|
| 20 | 18 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee+ff+cd+j+gj | 0.0158         | -857.1 | -762.4 | 0.7003         |
| 19 | 17 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee+ff+cd+j    | 0.0169         | -813.3 | -723.1 | 0.6792         |
| 18 | 17 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee+ff+cd      | 0.0170         | -809.4 | -723.7 | 0.6764         |
| 17 | 14 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee+ff         | 0.0199         | -703.5 | -622.3 | 0.6200         |
| 16 | 15 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce+ee            | 0.0216         | -650.6 | -573.9 | 0.5876         |
| 15 | 14 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj+ce               | 0.0217         | -647.4 | -575.2 | 0.5844         |
| 14 | 12 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd+aj                  | 0.0226         | -623.2 | -555.5 | 0.5679         |
| 13 | 12 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f+rd                     | 0.0233         | -604.0 | -540.8 | 0.5541         |
| 12 | 11 | ad+aa+de+bb+a+bi+ai+b+fg+dd+f                        | 0.0234         | -600.7 | -542.0 | 0.5505         |
| 11 | 11 | ad+aa+de+bb+a+bi+ai+b+fg+dd                          | 0.0246         | -569.1 | -515.0 | 0.5275         |
| 10 | 10 | ad+aa+de+bb+a+bi+ai+b+fg                             | 0.0300         | -435.6 | -386.0 | 0.4219         |
| 9  | 9  | ad+aa+de+bb+a+bi+ai+b                                | 0.0308         | -420.5 | -375.4 | 0.4070         |
| 8  | 8  | ad+aa+de+bb+a+bi+ai                                  | 0.0311         | -414.0 | -373.4 | 0.3995         |
| 7  | 6  | ad+aa+de+bb+a+bi                                     | 0.0341         | -353.7 | -317.6 | 0.3412         |
| 6  | 6  | ad+aa+de+bb+a  | 0.0340         | -355.4 | -323.9 | 0.3409         |
| 5  | 4  | ad+aa+de+bb  | 0.0387         | -269.8 | -242.7 | 0.2491         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

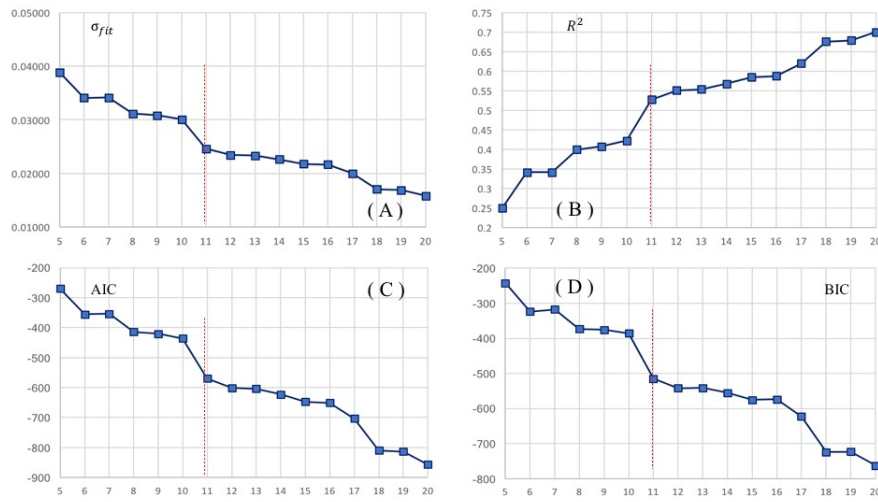


Figura A.47: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward-Forward.

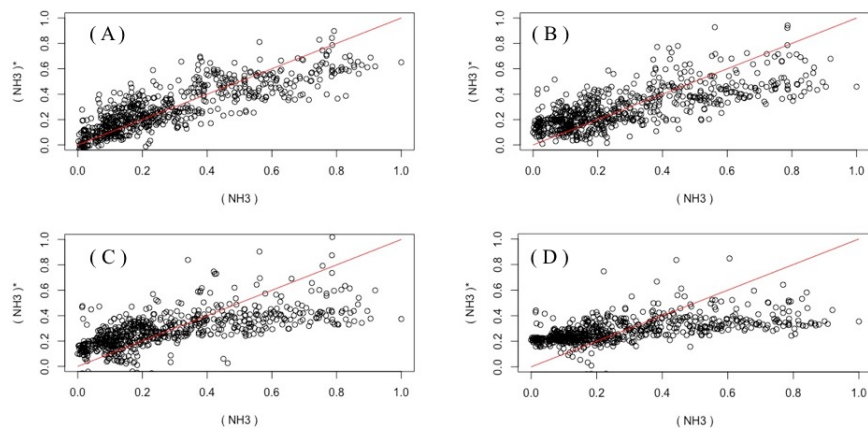


Figura A.48: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 15$ , (C) modelo deficitário com  $p = 14$  e (D) modelo extremo com  $p = 5$ .

A.1.7 NMVOCs

A.1.7.1 AIC-B

Tabela A.25: Evolução dos indicadores de desempenho do modelo AIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 11 | d+dd+a+af+ag+be+gg+g+dg+aa+ra+de+cd+f+rc+bc+rd+bb+bg | 0.00339        | -1892.5 | -1797.8 | 0.9012         |
| 19 | 14 | d+dd+a+af+ag+be+gg+g+dg+aa+ra+de+cd+f+rc+bc+rd+bb    | 0.00339        | -1893.5 | -1803.2 | 0.9011         |
| 18 | 14 | d+dd+a+af+ag+be+gg+g+dg+aa+ra+de+cd+f+rc+bc+rd       | 0.00339        | -1895.4 | -1809.7 | 0.9010         |
| 17 | 14 | d+dd+a+af+ag+be+gg+g+dg+aa+ra+de+cd+f+rc+bc          | 0.00342        | -1889.3 | -1808.1 | 0.8999         |
| 16 | 8  | d+dd+a+af+ag+be+gg+g+dg+aa+ra+de+cd+f+rc             | 0.00373        | -1831.9 | -1755.2 | 0.8906         |
| 15 | 8  | d+dd+a+af+ag+be+gg+g+dg+aa+ra+de+cd+f                | 0.00378        | -1824.7 | -1752.5 | 0.8891         |
| 14 | 6  | d+dd+a+af+ag+be+gg+g+dg+aa+ra+de+cd                  | 0.00380        | -1822.0 | -1754.4 | 0.8883         |
| 13 | 4  | d+dd+a+af+ag+be+gg+g+dg+aa+ra+de                     | 0.00393        | -1800.7 | -1737.5 | 0.8844         |
| 12 | 4  | d+dd+a+af+ag+be+gg+g+dg+aa+ra                        | 0.00393        | -1801.9 | -1743.3 | 0.8842         |
| 11 | 4  | d+dd+a+af+ag+be+gg+g+dg+aa                           | 0.00392        | -1803.9 | -1749.7 | 0.8842         |
| 10 | 7  | d+dd+a+af+ag+be+gg+g+dg                              | 0.00398        | -1795.3 | -1745.6 | 0.8824         |
| 9  | 6  | d+dd+a+af+ag+be+gg+g                                 | 0.00400        | -1792.9 | -1747.8 | 0.8816         |
| 8  | 6  | d+dd+a+af+ag+be+gg                                   | 0.00399        | -1794.5 | -1753.9 | 0.8816         |
| 7  | 6  | d+dd+a+af+ag+be                                      | 0.00399        | -1796.2 | -1760.1 | 0.8815         |
| 6  | 6  | d+dd+a+af+ag   | 0.00404        | -1789.2 | -1757.6 | 0.8799         |
| 5  | 5  | d+dd+a+af  | 0.00406        | -1786.4 | -1759.4 | 0.8790         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

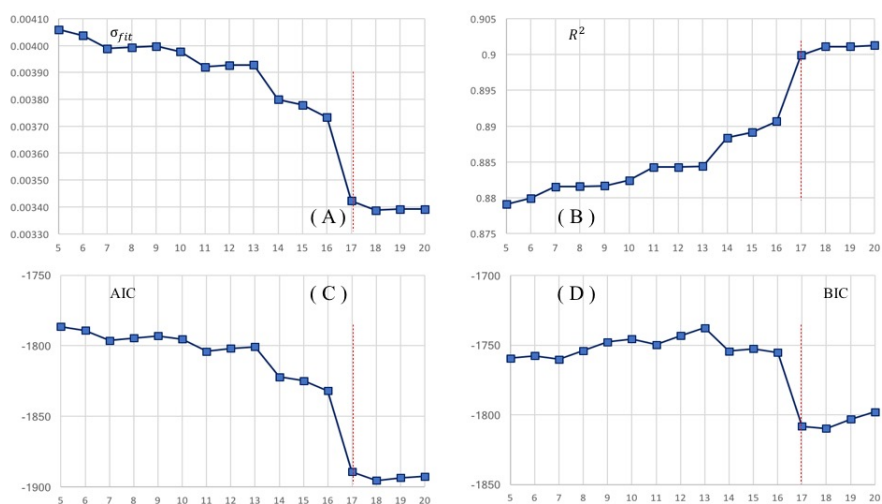


Figura A.49: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

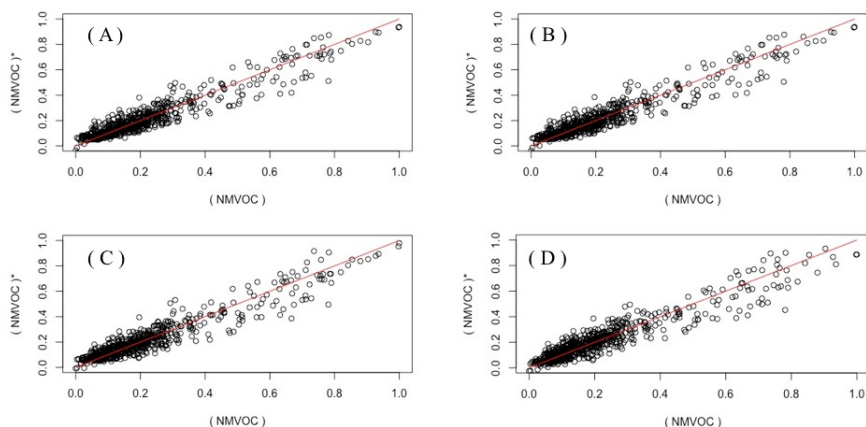


Figura A.50: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 17$ , (C) modelo deficitário com  $p = 16$  e (D) modelo extremo com  $p = 5$ .

### A.1.7.2 AIC-BF

Tabela A.26: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|--|----------------|---------|---------|----------------|
| 20 | 11 | d+dd+a+af+ag+be+aa+gg+g+dg+de+rc+cd+f+ra+bb+bc+rd+bg | 0.00339        | -1892.5 | -1797.8 | 0.9012         |
| 19 | 14 | d+dd+a+af+ag+be+aa+gg+g+dg+de+rc+cd+f+ra+bb+bc+rd    | 0.00339        | -1893.5 | -1803.2 | 0.9011         |
| 18 | 13 | d+dd+a+af+ag+be+aa+gg+g+dg+de+rc+cd+f+ra+bb+bc       | 0.00343        | -1887.5 | -1801.8 | 0.8999         |
| 17 | 8  | d+dd+a+af+ag+be+aa+gg+g+dg+de+rc+cd+f+ra+bb          | 0.00373        | -1830.7 | -1749.5 | 0.8907         |
| 16 | 8  | d+dd+a+af+ag+be+aa+gg+g+dg+de+rc+cd+f+ra             | 0.00373        | -1831.9 | -1755.2 | 0.8906         |
| 15 | 9  | d+dd+a+af+ag+be+aa+gg+g+dg+de+rc+cd+f                | 0.00373        | -1833.9 | -1761.7 | 0.8906         |
| 14 | 8  | d+dd+a+af+ag+be+aa+gg+g+dg+de+rc+cd                  | 0.00373        | -1834.4 | -1766.8 | 0.8904         |
| 13 | 6  | d+dd+a+af+ag+be+aa+gg+g+dg+de+rc                     | 0.00389        | -1806.9 | -1743.8 | 0.8854         |
| 12 | 5  | d+dd+a+af+ag+be+aa+gg+g+dg+de                        | 0.00392        | -1802.6 | -1743.9 | 0.8844         |
| 11 | 4  | d+dd+a+af+ag+be+aa+gg+g+dg                           | 0.00392        | -1803.9 | -1749.7 | 0.8842         |
| 10 | 4  | d+dd+a+af+ag+be+aa+gg+g                              | 0.00394        | -1802   | -1752.4 | 0.8836         |
| 9  | 4  | d+dd+a+af+ag+be+aa+gg                                | 0.00393        | -1803.8 | -1758.7 | 0.8835         |
| 8  | 3  | d+dd+a+af+ag+be+aa                                   | 0.00393        | -1805.8 | -1765.2 | 0.8835         |
| 7  | 6  | d+dd+a+af+ag+be                                      | 0.00399        | -1796.2 | -1760.1 | 0.8815         |
| 6  | 6  | d+dd+a+af+ag   | 0.00404        | -1789.2 | -1757.6 | 0.8799         |
| 5  | 5  | d+dd+a+af  | 0.00406        | -1786.4 | -1759.4 | 0.8790         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

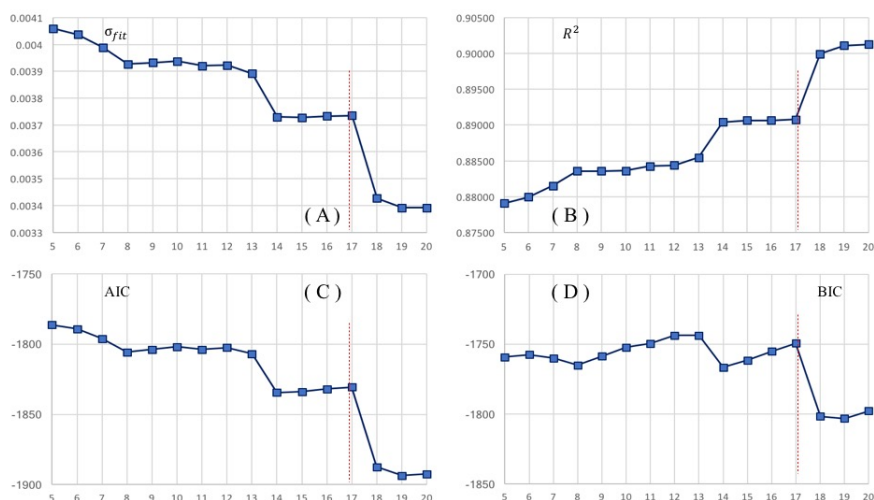


Figura A.51: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

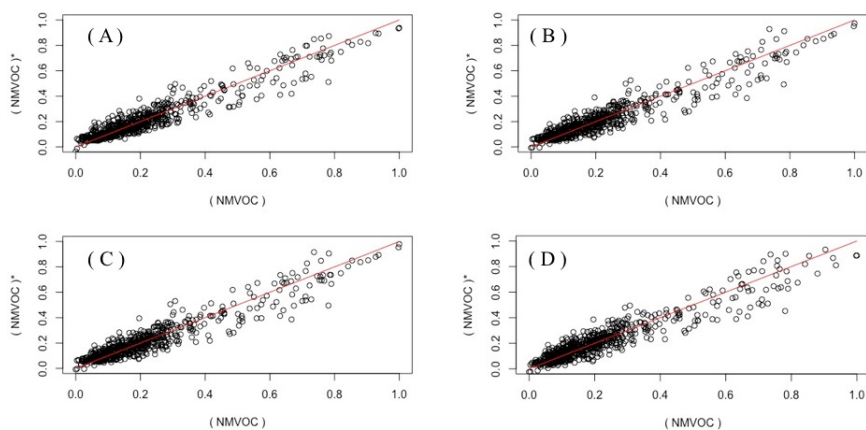


Figura A.52: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 17$ , (C) modelo deficitário com  $p = 16$  e (D) modelo extremo com  $p = 5$ .

A.1.7.3 BIC-B

Tabela A.27: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 13 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h+ef+rc+cj+gi | 0.00278        | -2021.6 | -1926.9 | 0.9185         |
| 19 | 17 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h+ef+rc+cj    | 0.00328        | -1915.9 | -1825.7 | 0.9043         |
| 18 | 15 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h+ef+rc       | 0.00341        | -1890.3 | -1804.6 | 0.9003         |
| 17 | 15 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h+ef          | 0.00347        | -1880.8 | -1799.6 | 0.8986         |
| 16 | 12 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h             | 0.00350        | -1874.8 | -1798.1 | 0.8974         |
| 15 | 12 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc               | 0.00351        | -1875.2 | -1803.0 | 0.8971         |
| 14 | 8  | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd                  | 0.00377        | -1826.6 | -1759.0 | 0.8891         |
| 13 | 4  | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f                     | 0.00393        | -1800.6 | -1737.5 | 0.8844         |
| 12 | 4  | d+a+af+g+dd+gg+ag+aa+dg+be+ra                       | 0.00393        | -1801.9 | -1743.3 | 0.8842         |
| 11 | 4  | d+a+af+g+dd+gg+ag+aa+dg+be                          | 0.00392        | -1803.9 | -1749.7 | 0.8842         |
| 10 | 4  | d+a+af+g+dd+gg+ag+aa+dg                             | 0.00393        | -1804.0 | -1754.3 | 0.8839         |
| 9  | 4  | d+a+af+g+dd+gg+ag+aa                                | 0.00394        | -1802.6 | -1757.5 | 0.8833         |
| 8  | 5  | d+a+af+g+dd+gg+ag                                   | 0.00405        | -1785.7 | -1745.1 | 0.8800         |
| 7  | 5  | d+a+af+g+dd+gg                                      | 0.00404        | -1787.6 | -1751.6 | 0.8800         |
| 6  | 6  | d+a+af+g+dd   | 0.00403        | -1789.6 | -1758.1 | 0.8800         |
| 5  | 4  | d+a+af+g  | 0.00413        | -1775.2 | -1748.1 | 0.8770         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

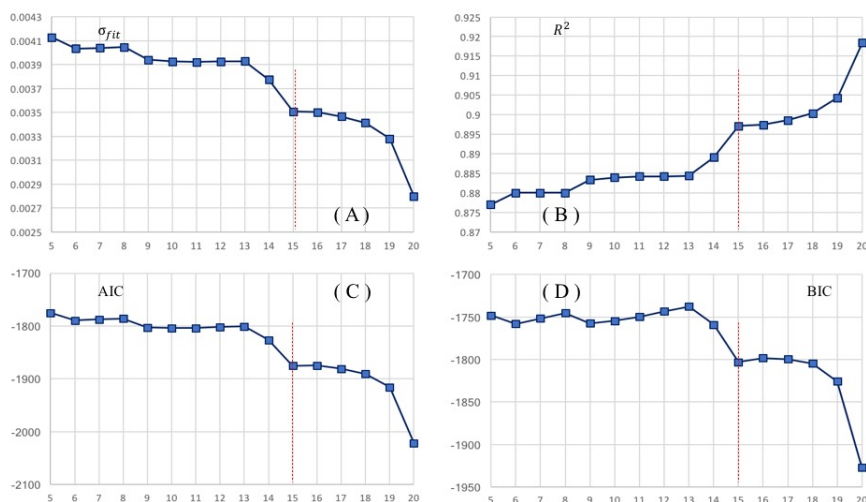


Figura A.53: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward.

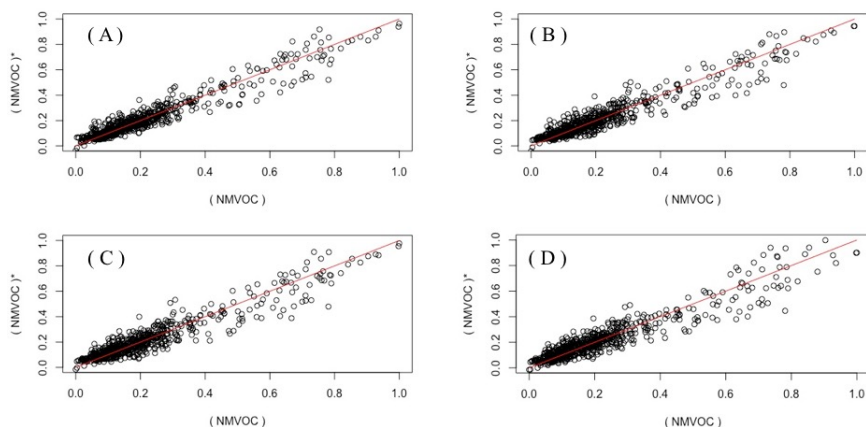


Figura A.54: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 15$ , (C) modelo deficitário com  $p = 14$  e (D) modelo extremo com  $p = 5$ .

#### A.1.7.4 BIC-BF

Tabela A.28: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo  | $\sigma_{fit}$ | AIC     | BIC     | R <sup>2</sup> |
|----|----|---|----------------|---------|---------|----------------|
| 20 | 13 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h+ef+rc+cj+gi | 0.00278        | -2021.6 | -1926.9 | 0.9185         |
| 19 | 17 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h+ef+rc+cj    | 0.00328        | -1915.9 | -1825.7 | 0.9043         |
| 18 | 15 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h+ef+rc       | 0.00341        | -1890.3 | -1804.6 | 0.9003         |
| 17 | 15 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h+ef          | 0.00347        | -1880.8 | -1799.6 | 0.8986         |
| 16 | 12 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc+h             | 0.00350        | -1874.8 | -1798.1 | 0.8974         |
| 15 | 12 | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd+bc               | 0.00351        | -1875.2 | -1803.0 | 0.8971         |
| 14 | 8  | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f+cd                  | 0.00377        | -1826.6 | -1759.0 | 0.8891         |
| 13 | 4  | d+a+af+g+dd+gg+ag+aa+dg+be+ra+f                     | 0.00393        | -1800.6 | -1737.5 | 0.8844         |
| 12 | 4  | d+a+af+g+dd+gg+ag+aa+dg+be+ra                       | 0.00393        | -1801.9 | -1743.3 | 0.8842         |
| 11 | 4  | d+a+af+g+dd+gg+ag+aa+dg+be                          | 0.00392        | -1803.9 | -1749.7 | 0.8842         |
| 10 | 4  | d+a+af+g+dd+gg+ag+aa+dg                             | 0.00393        | -1804.0 | -1754.3 | 0.8839         |
| 9  | 4  | d+a+af+g+dd+gg+ag+aa                                | 0.00394        | -1802.6 | -1757.5 | 0.8833         |
| 8  | 5  | d+a+af+g+dd+gg+ag                                   | 0.00405        | -1785.7 | -1745.1 | 0.8800         |
| 7  | 5  | d+a+af+g+dd+gg                                      | 0.00404        | -1787.6 | -1751.6 | 0.8800         |
| 6  | 6  | d+a+af+g+dd   | 0.00403        | -1789.6 | -1758.1 | 0.8800         |
| 5  | 4  | d+a+af+g  | 0.00413        | -1775.2 | -1748.1 | 0.8770         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;



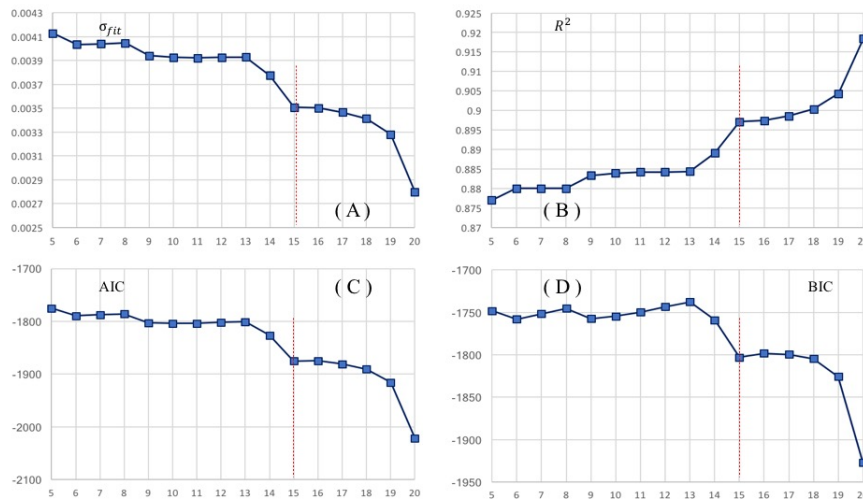


Figura A.55: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward-Forward.

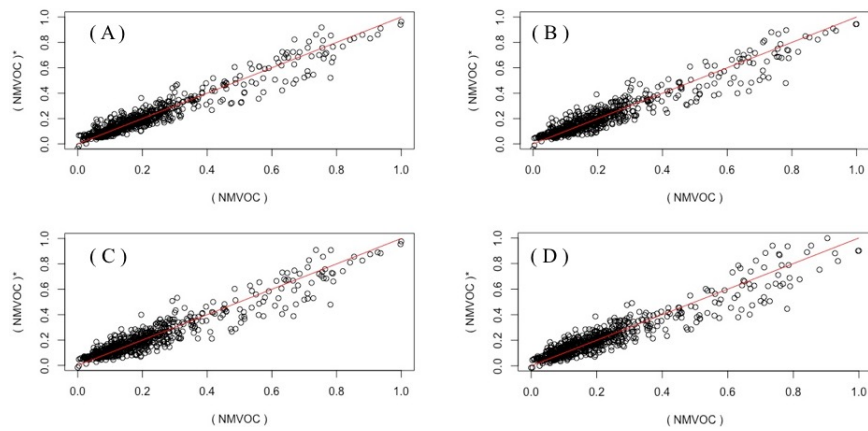


Figura A.56: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 15$ , (C) modelo deficitário com  $p = 14$  e (D) modelo extremo com  $p = 5$ .

**A.1.8 PANs**

**A.1.8.1 AIC-B**

Tabela A.29: Evolução dos indicadores de desempenho do modelo AIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC    | BIC    | R <sup>2</sup> |
|----|----|--|----------------|--------|--------|----------------|
| 20 | 7  | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a+gi+fg+eh+ri | 0.0177         | -779.1 | -684.4 | 0.5943         |
| 19 | 9  | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a+gi+fg+eh    | 0.0177         | -780.6 | -690.4 | 0.5940         |
| 18 | 10 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a+gi+fg       | 0.0177         | -782.2 | -696.5 | 0.5938         |
| 17 | 10 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a+gi          | 0.0177         | -784.0 | -702.8 | 0.5936         |
| 16 | 11 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a             | 0.0177         | -785.9 | -709.2 | 0.5936         |
| 15 | 10 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg               | 0.0195         | -720.3 | -648.1 | 0.5506         |
| 14 | 10 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh                  | 0.0195         | -722.2 | -654.5 | 0.5505         |
| 13 | 8  | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg                     | 0.0210         | -672.8 | -609.6 | 0.5148         |
| 12 | 9  | ae+be+bi+ag+ab+i+g+bd+de+b+dg                        | 0.0210         | -672.8 | -614.2 | 0.5133         |
| 11 | 9  | ae+be+bi+ag+ab+i+g+bd+de+b                           | 0.0212         | -666.8 | -612.7 | 0.5075         |
| 10 | 9  | ae+be+bi+ag+ab+i+g+bd+de                             | 0.0248         | -565.3 | -515.7 | 0.4255         |
| 9  | 9  | ae+be+bi+ag+ab+i+g+bd                                | 0.0255         | -546.5 | -501.4 | 0.4074         |
| 8  | 6  | ae+be+bi+ag+ab+i+g                                   | 0.0297         | -444.3 | -403.7 | 0.3081         |
| 7  | 5  | ae+be+bi+ag+ab+i                                     | 0.0325         | -386.6 | -350.5 | 0.2437         |
| 6  | 4  | ae+be+bi+ag+ab                                       | 0.0340         | -356.1 | -324.5 | 0.2063         |
| 5  | 3  | ae+be+bi+ag  | 0.0343         | -352.0 | -324.9 | 0.1990         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

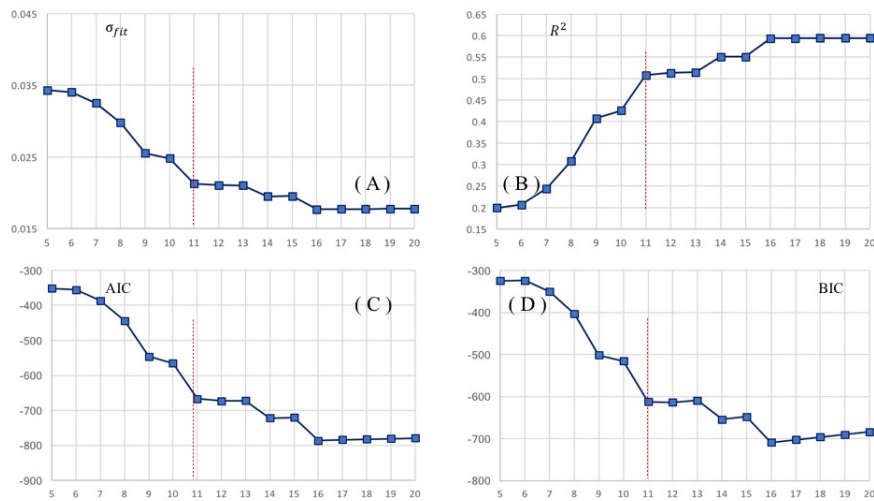


Figura A.57: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward.

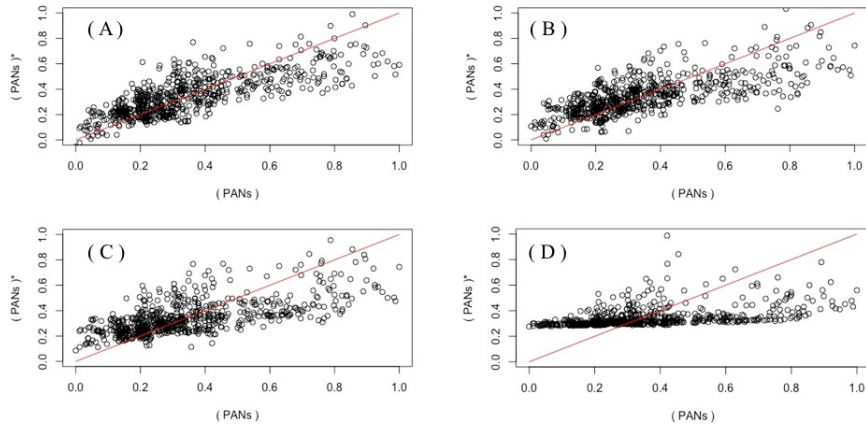


Figura A.58: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 11$ , (C) modelo deficitário com  $p = 10$  e (D) modelo extremo com  $p = 5$ .

**A.1.8.2 AIC-BF**

Tabela A.30: Evolução dos indicadores de desempenho do modelo AIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC    | BIC    | R <sup>2</sup> |
|----|----|--|----------------|--------|--------|----------------|
| 20 | 7  | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a+gi+fg+eh+ri | 0.0177         | -779.1 | -684.4 | 0.5943         |
| 19 | 9  | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a+gi+fg+eh    | 0.0177         | -780.6 | -690.4 | 0.5940         |
| 18 | 10 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a+gi+fg       | 0.0177         | -782.2 | -696.5 | 0.5938         |
| 17 | 10 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a+gi          | 0.0177         | -784.0 | -702.8 | 0.5936         |
| 16 | 11 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg+a             | 0.0177         | -785.9 | -709.2 | 0.5936         |
| 15 | 10 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh+bg               | 0.0195         | -720.3 | -648.1 | 0.5506         |
| 14 | 10 | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg+dh                  | 0.0195         | -722.2 | -654.5 | 0.5505         |
| 13 | 8  | ae+be+bi+ag+ab+i+g+bd+de+b+dg+eg                     | 0.0210         | -672.8 | -609.6 | 0.5148         |
| 12 | 9  | ae+be+bi+ag+ab+i+g+bd+de+b+dg                        | 0.0210         | -672.8 | -614.2 | 0.5133         |
| 11 | 9  | ae+be+bi+ag+ab+i+g+bd+de+b                           | 0.0212         | -666.8 | -612.7 | 0.5075         |
| 10 | 9  | ae+be+bi+ag+ab+i+g+bd+de                             | 0.0248         | -565.3 | -515.7 | 0.4255         |
| 9  | 9  | ae+be+bi+ag+ab+i+g+bd                                | 0.0255         | -546.5 | -501.4 | 0.4074         |
| 8  | 6  | ae+be+bi+ag+ab+i+g                                   | 0.0297         | -444.3 | -403.7 | 0.3081         |
| 7  | 5  | ae+be+bi+ag+ab+i                                     | 0.0325         | -386.6 | -350.5 | 0.2437         |
| 6  | 4  | ae+be+bi+ag+ab                                       | 0.0340         | -356.1 | -324.5 | 0.2063         |
| 5  | 3  | ae+be+bi+ag  | 0.0343         | -352.0 | -324.9 | 0.1990         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

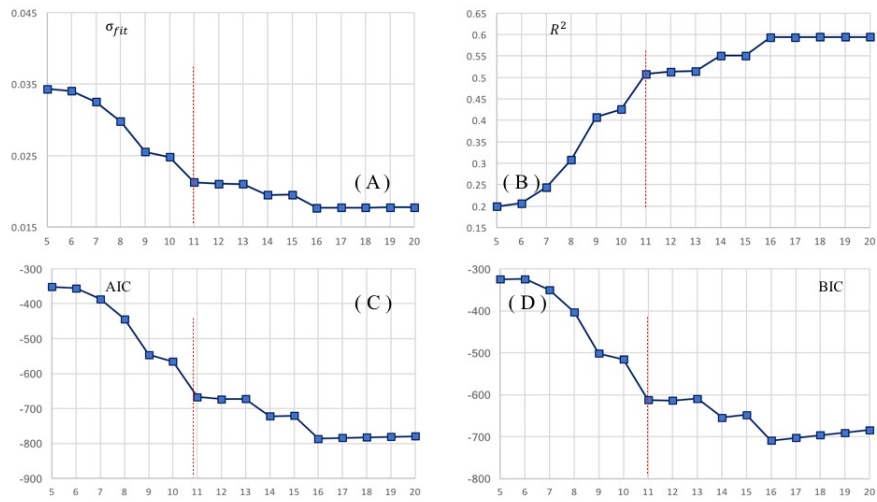


Figura A.59: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério AIC-Backward-Forward.

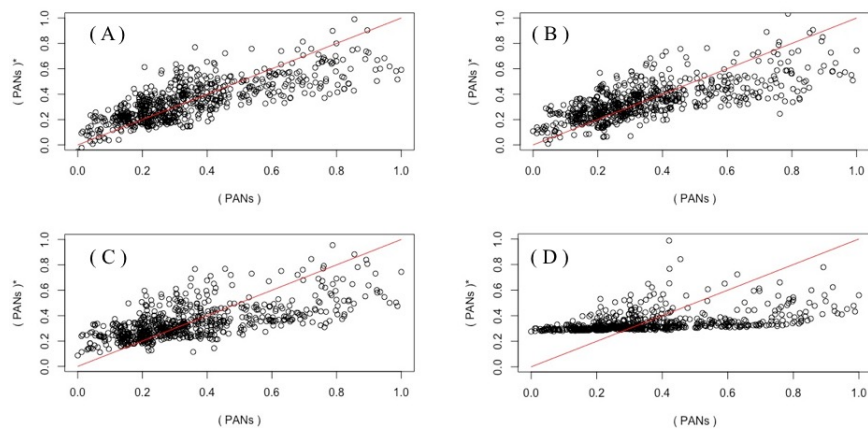


Figura A.60: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 11$ , (C) modelo deficitário com  $p = 10$  e (D) modelo extremo com  $p = 5$ .

**A.1.8.3 BIC-B**

Tabela A.31: Evolução dos indicadores de desempenho do modelo BIC-B em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC    | BIC    | R <sup>2</sup> |
|----|----|--|----------------|--------|--------|----------------|
| 20 | 19 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri+rg+hh+dh+ce | 0.0134         | -967.4 | -872.6 | 0.6934         |
| 19 | 12 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri+rg+hh+dh    | 0.0166         | -825.1 | -734.9 | 0.6200         |
| 18 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri+rg+hh       | 0.0169         | -814.5 | -728.8 | 0.6128         |
| 17 | 12 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri+rg          | 0.0181         | -768.1 | -686.9 | 0.5839         |
| 16 | 10 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri             | 0.0204         | -690.7 | -614.1 | 0.5318         |
| 15 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg                | 0.0207         | -679.9 | -607.7 | 0.5227         |
| 14 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de                   | 0.0207         | -681.9 | -614.2 | 0.5227         |
| 13 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a                      | 0.0214         | -658.9 | -595.8 | 0.5046         |
| 12 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg                        | 0.0226         | -623.3 | -564.7 | 0.4761         |
| 11 | 10 | ae+bi+be+b+ab+i+ag+g+eg+gi                           | 0.0229         | -615.9 | -561.8 | 0.4688         |
| 10 | 9  | ae+bi+be+b+ab+i+ag+g+eg                              | 0.0230         | -614.9 | -565.3 | 0.4664         |
| 9  | 9  | ae+bi+be+b+ab+i+ag                                   | 0.0238         | -593.0 | -547.9 | 0.4471         |
| 8  | 8  | ae+bi+be+b+ab+i                                      | 0.0249         | -563.6 | -523.0 | 0.4206         |
| 7  | 7  | ae+bi+be+b+ab  | 0.0257         | -543.7 | -507.7 | 0.4015         |
| 6  | 5  | ae+bi+be+b   | 0.0363         | -312.7 | -281.2 | 0.1534         |
| 5  | 5  | ae+bi+be   | 0.0375         | -292.1 | -265.0 | 0.1244         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

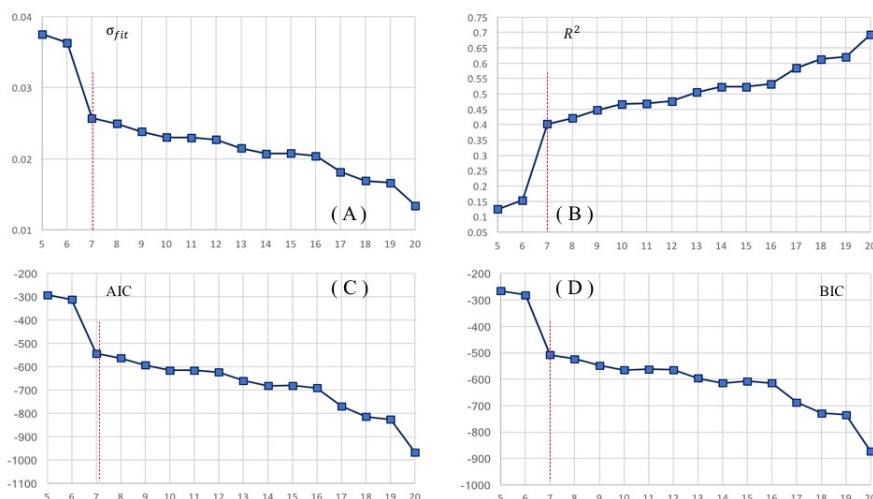


Figura A.61: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward.

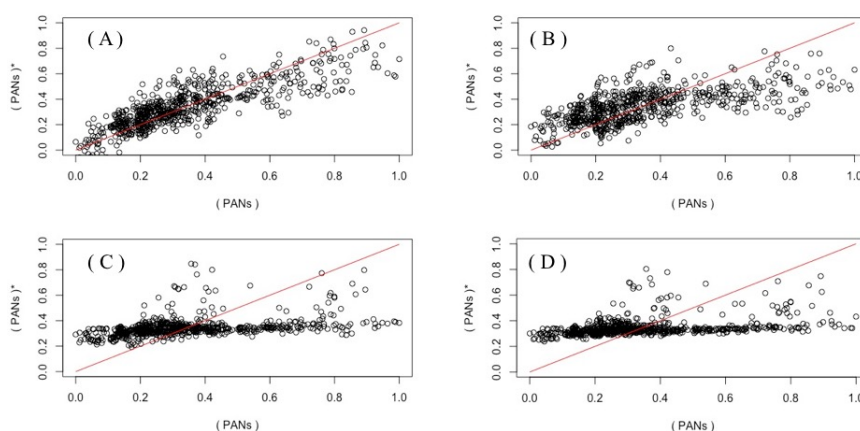


Figura A.62: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 7$ , (C) modelo deficitário com  $p = 6$  e (D) modelo extremo com  $p = 5$ .

#### A.1.8.4 BIC-BF

Tabela A.32: Evolução dos indicadores de desempenho do modelo BIC-BF em função do conjunto de parâmetros utilizado.

| p  | p* | Modelo   | $\sigma_{fit}$ | AIC    | BIC    | R <sup>2</sup> |
|----|----|--|----------------|--------|--------|----------------|
| 20 | 19 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri+rg+hh+dh+ce | 0.0134         | -967.4 | -872.6 | 0.6934         |
| 19 | 12 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri+rg+hh+dh    | 0.0166         | -825.1 | -734.9 | 0.6200         |
| 18 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri+rg+hh       | 0.0169         | -814.5 | -728.8 | 0.6128         |
| 17 | 12 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri+rg          | 0.0181         | -768.1 | -686.9 | 0.5839         |
| 16 | 10 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg+ri             | 0.0204         | -690.7 | -614.1 | 0.5318         |
| 15 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de+bg                | 0.0207         | -679.9 | -607.7 | 0.5227         |
| 14 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a+de                   | 0.0207         | -681.9 | -614.2 | 0.5227         |
| 13 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg+a                      | 0.0214         | -658.9 | -595.8 | 0.5046         |
| 12 | 11 | ae+bi+be+b+ab+i+ag+g+eg+gi+fg                        | 0.0226         | -623.3 | -564.7 | 0.4761         |
| 11 | 10 | ae+bi+be+b+ab+i+ag+g+eg+gi                           | 0.0229         | -615.9 | -561.8 | 0.4688         |
| 10 | 9  | ae+bi+be+b+ab+i+ag+g+eg                              | 0.0230         | -614.9 | -565.3 | 0.4664         |
| 9  | 9  | ae+bi+be+b+ab+i+ag                                   | 0.0238         | -593.0 | -547.9 | 0.4471         |
| 8  | 8  | ae+bi+be+b+ab+i+ag                                   | 0.0249         | -563.6 | -523.0 | 0.4206         |
| 7  | 7  | ae+bi+be+b+ab+i                                      | 0.0257         | -543.7 | -507.7 | 0.4015         |
| 6  | 5  | ae+bi+be+b+ab  | 0.0363         | -312.7 | -281.2 | 0.1534         |
| 5  | 5  | ae+bi+be+b   | 0.0375         | -292.1 | -265.0 | 0.1244         |

p - número de parâmetros do modelo; p\* - número de parâmetros com significado estatístico;

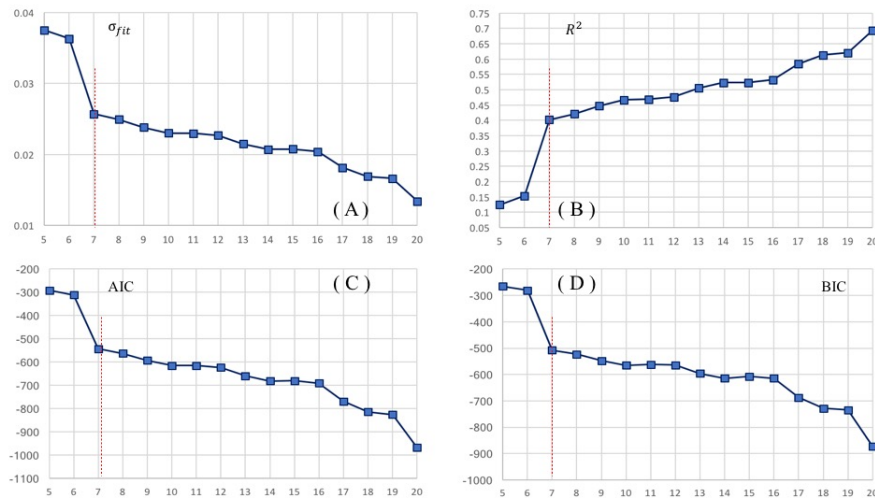


Figura A.63: Evolução da capacidade de ajuste do modelo em função do número de parâmetros utilizados: (A) desvio padrão do ajuste, (B) coeficiente de determinação, (C) AIC e (D) BIC para o sub conjunto de parâmetros refinados através do critério BIC-Backward-Forward.

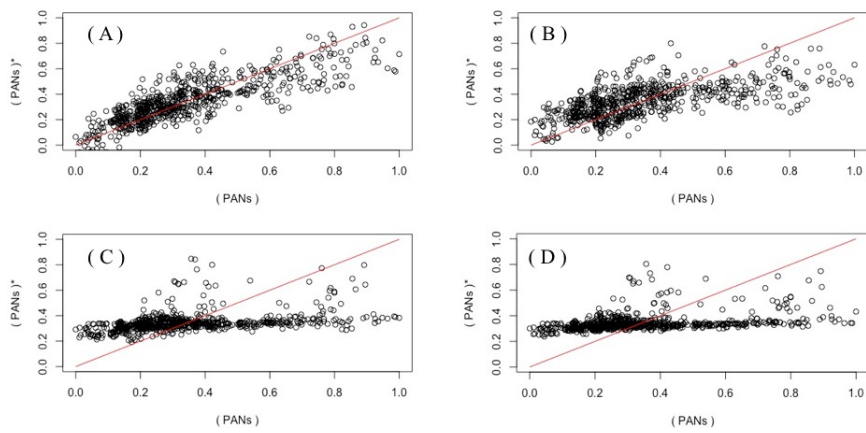


Figura A.64: Evolução da recuperação de informação do modelo em função do número de parâmetros utilizados: (A) modelo mais completo com  $p = 20$ , (B) modelo parsimonioso com  $p = 7$ , (C) modelo de citário com  $p = 6$  e (D) modelo extremo com  $p = 5$ .