

# MOS – “Model Output Statistics” – Aplicação da metodologia para correcção das previsões de vento obtidas para Portugal Continental

R. Marujo<sup>(1)</sup>, P. Costa<sup>(1)</sup>, A. Estanqueiro<sup>(1)</sup>, C. Pires<sup>(2)</sup>

[raquel.marujo@lneg.pt](mailto:raquel.marujo@lneg.pt), [paulo.costa@lneg.pt](mailto:paulo.costa@lneg.pt), [ana.estanqueiro@ineti.pt](mailto:ana.estanqueiro@ineti.pt), [capires@fc.ul.pt](mailto:capires@fc.ul.pt)

<sup>(1)</sup> LNEG (ex. INETI) – Laboratório Nacional de Energia e Geologia

<sup>(2)</sup> FCUL – Faculdade de Ciências da Universidade de Lisboa

## Resumo

A previsão do vento com exactidão é actualmente uma ferramenta necessário para a análise do potencial e estimativa de produção de energia eólica. Apesar do progresso significativo nos modelos de previsão meteorológica, os modelos apresentam erros sistemáticos que podem ser corrigidos com a aplicação de modelos estatísticos, lineares ou não, a partir das estatísticas acumuladas de um certo modelo de previsão. Esses modelos correctivos têm o nome genérico de MOS (Model Output Statistics) que têm como base o uso de regressões multilíneas. Neste estudo, realizaram-se previsões MOS para as 24, 48 e 72 horas, para as componentes horizontais do vento, assim como, para o prazo de previsão de 36 horas, nas 61 estações da rede do Instituto de Meteorologia Português. Os resultados mostram uma melhoria do erro médio quadrático na ordem de 80% na previsão, sendo menos significativa com o aumento do prazo de previsão.

## Introdução

Durante a última década assistiu-se em Portugal a um forte crescimento na instalação de Parques Eólicos, resultado de um enorme investimento na área das Energias Renováveis em geral, e do elevado potencial eólico do nosso país. Para a instalação de qualquer sistema de energia é necessário realizar um estudo minucioso acerca das grandezas que o definem. No caso da energia eólica, o escoamento atmosférico constitui a sua fonte primária, como tal, o estudo foca as suas características gerais e locais. A previsão do vento com exactidão é uma ferramenta necessária para a estimativa da produção de energia eólica, uma vez que a potência depende linearmente do cubo da intensidade do vento.

A evolução no progresso dos modelos de previsão tem sido bastante satisfatória ao longo das últimas décadas. O output das simulações numéricas são previsões dos campos das variáveis meteorológicas em diferentes níveis numa grelha específica. Nestas previsões existem erros sistemáticos que não dependem somente das falhas de parametrização física dos modelos dinâmicos, mas também da topografia local, da resolução do modelo e dos

erros de interpolação das estações para os pontos de grelha do modelo entre outras causas. Para interpretar estes erros e caracterizá-los existe a possibilidade de implementar um modelo estatístico que use como variáveis de entrada as simulações do modelo numérico – MOS.

## Metodologia

Os modelos de previsão estatísticos têm por base equações de regressões lineares múltiplas, onde a variável dependente é expressa em função das quantidades que contribuem para a sua variação [1 e 2].

O MOS é um exemplo de um modelo estatístico que estabelece relações entre as previsões do modelo numérico e as observações. O caso mais simplificado deste método é a equação da regressão linear simples, que estabelece a relação entre um único predictor e um predictando.

### Regressão Multilinear

Nos casos onde se considera  $k$  predictores, as equações de previsão são obtidas através de regressões lineares múltiplas. A equação de previsão vem:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (1)$$

onde  $b_0$  é denominado a constante da regressão e os coeficientes da regressão são os termos  $b_i$  ( $i=1, \dots, k$ ), correspondentes aos  $k$  predictores lineares.

### Análise da qualidade da previsão

As medidas estatísticas como o viés (BE), a raiz do erro médio quadrático (RMSE), o erro médio quadrático (MSE) e a correlação de Pearson ( $r$ ), podem ser usados para quantificar a capacidade do modelo.

$$BE(x) = \frac{1}{N} \sum_{i=1}^N (x_i^p - x_i^o) \quad (2)$$

$$MSE(x) = \frac{1}{N} \sum_{i=1}^N (x_i^p - x_i^o)^2 \quad (3)$$

$$RMSE(x) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i^p - x_i^o)^2} \quad (4)$$

$$r = \frac{cov(x,y)}{\sqrt{var(x)var(y)}} \quad (5)$$

O score (SC- MESS – Mean square error skill score), também, irá ser uma medida adoptada para a avaliação da qualidade do modelo estatístico obtido para cada caso de estudo e variável.

$$SC = \frac{MSE_{MM5} - MSE_{MOS}}{MSE_{MM5}} \quad (6)$$

### Construção do modelo estatístico MOS

Para a construção das equações de regressão foram pré-seleccionados os predictores para cada caso de estudo, obtendo as equações com uma regressão forward stepwise e aferidas em validação cruzada [1 e 2].

- *Stepwise regression*

A filtragem da regressão consiste em seleccionar, de entre o conjunto de potenciais predictores, um subconjunto de predictores estatisticamente robustos. A metodologia da regressão stepwise (stepwise regression) consiste em seleccionar este sub-conjunto através da exclusão dos piores ou melhores predictores.

O método escolhido para esta selecção foi o forward stepwise regression, que consiste em adicionar, sucessivamente, o melhor predictor disponível, na elaboração de cada equação de regressão, em cada estação e variável de previsão.

- *Validação Cruzada*

O MOS será aferido em validação cruzada, onde as equações de regressão utilizadas para previsão meteorológica são testadas numa amostra de dados independentes da amostra de calibração do modelo estatístico.

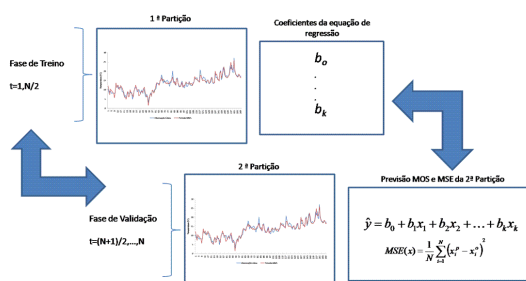


Fig.1.- Diagrama do método aplicado para o processamento da previsão (regressão multilinear aferido em validação cruzada).

A figura1 ilustra o método de validação aplicado. A validação cruzada é feita usando os conjuntos de dados de desenvolvimento do tamanho  $N/2$  (6 meses), e os conjuntos de verificação contendo a restante observação do predictando, onde  $N$  é o número total de dados da série. O modelo de regressão é recalculado para cada uma das duas partições (fase de treino e fase de validação). Uma vez obtidas as previsões, calcula-se o erro médio quadrático (MSE) de previsão, obtido pela média dos quadrados das diferenças entre observações e previsões. O processo de validação cruzada usa todas as  $N$  observações do predictando para estimar o MSE, permitindo que cada observação seja tratada de cada vez, como um dado independente.

### Dados Disponíveis

Para este trabalho foram disponibilizadas para o ano completo de 2005, previsões diárias do vento e das respectivas componentes horizontais com um horizonte temporal (prazo de previsão) até 72 horas, oriundas do modelo MM5 (The Fifth-Generation NCAR / Penn State Mesoscale Model (MM5) [3]), instalado na Faculdade de Ciências da Universidade de Lisboa a correr operacionalmente com uma malha de resolução espacial de 6km x 6km [4]. Foram processadas, com intervalos de tempo de 30 em 30 minutos, séries anuais de vento interpoladas para os pontos das estações da rede do Instituto de Meteorologia.

A base de dados das observações foi disponibilizada em intervalos de 10 minutos para o ano de 2005. Para a construção das equações de regressão do modelo estatístico MOS foram construídas séries, com espaçamento temporal de 30 minutos, para as variáveis do vento e as respectivas componente zonal ( $u$ ) e meridional ( $v$ ).

### Casos de estudo

O estudo inicial (caso 1) teve como base a construção das equações de regressão do MOS para uma previsão pontual a uma hora específica do prazo de previsão (24, 48 e 72 horas), para as componentes horizontais do vento, nas 61 estações da rede do Instituto de Meteorologia Português, para o ano de 2005 [5].

Tabela 1 – Predictores seleccionados para o MOS (24h).

Casos	Predictores
A	P(+24)
B	P(+24) O(t)
C	P(+24) O(t) O(t-12)
D	P(+24) O(t) O(t-12) P(+12)
E	P(+24) O(t) O(t-12) P(+12) P(+36)

Foram obtidas dois tipos de equações de regressão, para cinco casos experimentais, em que as principais diferenças são a selecção e o número de predictores. Para os cinco casos foi construído o

modelo com selecção de predictores, através de uma regressão stepwise, e sem selecção, ou seja seguindo a ordem apresentada na tabela 1. Todos os modelos foram aferidos em validação cruzada.

Os predictores pré-seleccionados foram previsões do modelo MM5 e observações, disponíveis no lançamento da previsão (P(t+24) O(t) O(t-12) P(t+12) P(t+36)) (figura 2).

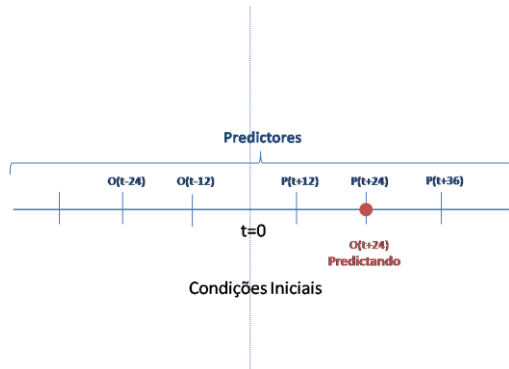


Fig.2.- Representação espacial e temporal dos predictores pré-seleccionados para o casos do estudo inicial.

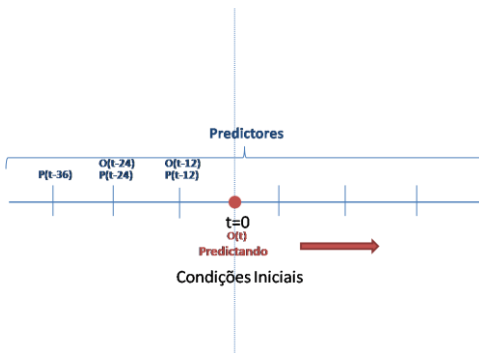


Fig.3.- Representação espacial e temporal dos predictores pré-seleccionados para o caso de estudo 2.

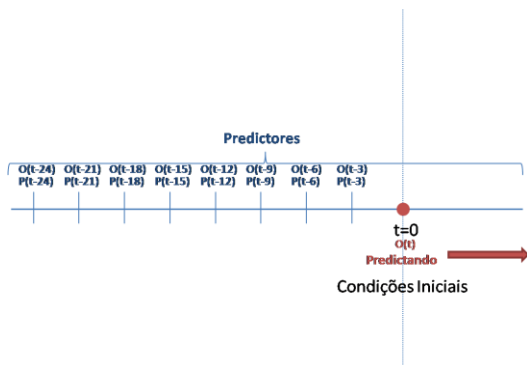


Fig.4.- Representação espacial e temporal dos predictores pré-seleccionados para o caso de estudo 3.

Aplicando o mesmo método foram realizados dois estudos detalhados e direccionados para a energia eólica, através da previsão estatística para o prazo de 36 horas de previsão para a variável do vento e as respectivas componentes horizontais. Os predictores utilizados foram pré-seleccionados de acordo com os dados disponíveis no lançamento da previsão. Deste modo, foram seleccionados como

predictores as previsões e observações do dia anterior à realização da previsão com espaçamento temporal de 12h (figura 3) (6 predictores - P(t-12) P(t-24) P(t-36) O(t-12) O(t-24)) e de 3h (figura 4) (17 predictores - P(t-3) P(t-6) P(t-9) P(t-12) P(t-15) P(t-18) P(t-21) P(t-24) O(t-3) O(t-6) O(t-9) O(t-12) O(t-15) O(t-18) O(t-21) O(t-24)).

## Resultados

### Caso 1

A base de partida para a construção das equações de regressão do modelo estatístico, que visa minimizar os desvios médios e melhorar a previsão local dos parâmetros em estudo, é a análise da qualidade dos dados de input (dados do modelo numérico MM5). As figuras 5, 6 e 7 mostram os resultados das médias anuais para a previsão a 24 horas da correlação, do RMSE e do viés.

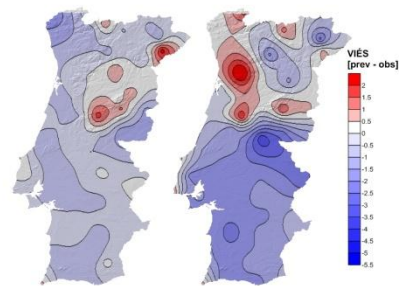


Fig. 5. – Viés (BE) entre as previsões do MM5 e as observações da rede IM. (componente u do vento; componente v do vento).

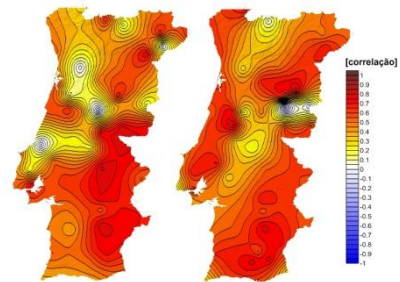


Fig. 6. – Correlação (r) entre as previsões do MM5 e as observações da rede IM. (componente u do vento; componente v do vento).

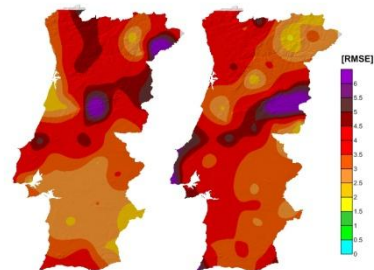


Fig. 7. – Raiz do erro médio quadrático (RMSE) entre as previsões do MM5 e as observações da rede IM. (componente u do vento; componente v do vento).

Em geral, os resultados demonstram que o modelo subestima as componentes do vento, excepto na zona da Serra da Estrela. Contudo, o modelo vai sobreavaliar o vento nas zonas em altitude, principalmente na componente v do vento.

Na tabela 2 estão apresentados os resultados estatísticos para os cinco casos do estudo inicial para a previsão a 24, 48 e 72 horas. No estudo a 48 e 72 horas apenas foi desenvolvido o modelo MOS para os dois primeiros casos (caso A e B), com selecção da ordem dos predictores para o caso B.

Tabela 2- Resumo dos parâmetros estatísticos de avaliação da qualidade da previsão dos modelos MOS e MM5, para a temperatura e cada componente horizontal do vento.

Casos	Hora	Seleccção	Variáveis	Parâmetros Estatísticos				
				BE	r	RMSE	MSE	SC [%]
MM5	24	---	u	-0.25	0.49	3.83	14.66	----
			v	-0.67	0.51	4.07	16.55	
	48	---	u	-0.33	0.41	3.98	15.82	
			v	0.50	0.49	3.54	12.53	
	72	---	u	-0.08	0.40	3.77	14.18	
			v	0.37	0.46	3.62	13.13	
MOSA	24	---	u	-0.02	0.50	1.77	3.14	79
			v	0.02	0.53	1.69	2.87	83
	48	---	u	-0.06	0.41	1.88	3.54	78
			v	0.003	0.49	1.75	3.05	76
	72	---	u	-0.04	0.40	1.90	3.60	75
			v	0.019	0.46	1.78	3.19	76
MOSB	24	sim	u	-0.02	0.50	1.77	3.14	79
		não	v	0.02	0.53	1.69	2.87	83
	48	sim	u	-0.02	0.37	1.98	3.91	75
		não	v	0.04	0.46	1.50	2.24	82
	72	sim	u	-0.03	0.39	1.91	3.65	74
		não	v	0.04	0.46	1.79	3.22	75
MOSC	24	sim	u	-0.03	0.58	1.66	2.76	81
			v	0.01	0.59	1.56	2.43	85
		não	u	0.034	0.54	1.75	3.06	79
			v	0.047	0.59	1.61	2.60	84
MOSD	24	sim	u	-0.042	0.59	1.64	2.68	82
			v	0.02	0.61	1.53	2.33	86
		não	u	-0.079	0.51	1.85	3.43	77
			v	0.098	0.57	1.78	3.17	81
MOSE	24	sim	u	-0.03	0.62	1.58	2.49	82
			v	0.03	0.64	1.48	2.19	86
		não	u	-0.09	0.50	1.89	3.57	76
			v	0.115	0.54	1.80	3.25	80

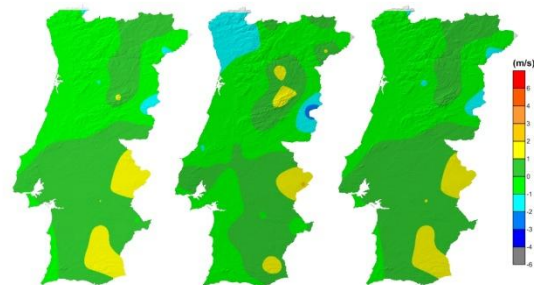
Em primeira análise, os melhores resultados obtidos são verificados no caso E, para a previsão de 24 horas e com selecção de predictores (figura 8). Estes resultados indicam uma melhoria na previsão das variáveis na ordem de 85%, diminuindo a qualidade com o aumento do prazo de previsão, devido aos dados de input associados às previsões do modelo MM5. Em geral, para todos os casos de estudo, a selecção de predictores, através da técnica da regressão forward stepwise apresenta sempre melhores resultados. Estes resultados mostram uma diminuição do erro médio quadrático e do viés, e uma ligeira melhoria da correlacção (aumento de 10% - caso E).

Os resultados do modelo para os casos onde não se seleccionou a ordem dos predictores é menos satisfatória. Estes resultados mostram que a partir da adição do quarto e quinto predictor, o modelo deixa de apresentar melhoria nos resultados.

Uma segunda análise, permite-nos verificar que no caso A, onde o modelo de regressão apresenta um único predictor (a previsão das 24 horas – P(t+24h)), a previsão melhora cerca de 80% para as componentes do vento. Com a adição de um

novo predictor na construção do modelo os valores das previsões melhoram cerca de 2 %, em relação ao caso A. Com a adição contínua e sucessiva dos predictores, a previsão do modelo sofre apenas ligeiras melhorias, inferiores ao verificado na introdução do segundo predictor – caso B. A introdução sucessiva de mais predictores apenas permite uma refinação do modelo.

**Componente Zonal do Vento (24 horas)**  
(Observação, Previsão MM5 e Previsão MOS)



**Componente Meridional do Vento (24 horas)**  
(Observação, Previsão MM5 e Previsão MOS)



Fig.8.- Resultados, da observação, previsão MM5 e MOS para as 24 horas, da média das componentes zonal e meridional do vento [4].

### Caso 2 e 3

Nas figuras 9,10 e 11 estão representados os mapas da distribuição das medidas estatísticas avaliadas para a análise da qualidade da previsão do modelo MM5 para o prazo de 36 horas de previsão. Os valores apresentados são uma média anual do viés, da correlação e do RMSE para as três variáveis (vento e as respectivas componentes horizontais).

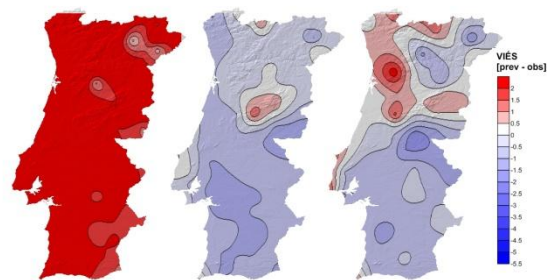


Fig. 9. – Viés (BE) entre as previsões do MM5 e as observações da rede IM. (componente u do vento; componente v do vento).

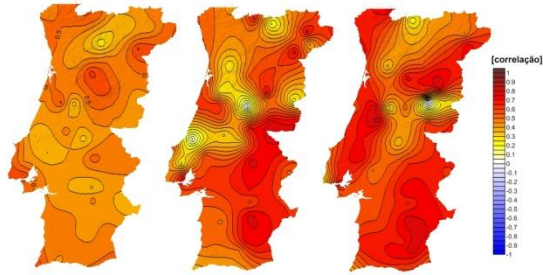


Fig. 10. – Correlação (r) entre as previsões do MM5 e as observações da rede IM. (componente u do vento; componente v do vento).

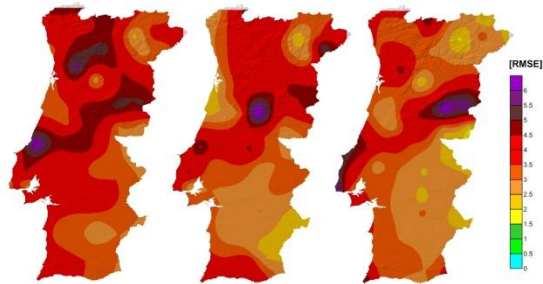


Fig. 11. – Raiz do erro médio quadrático (RMSE) entre as previsões do MM5 e as observações da rede IM. (componente u do vento; componente v do vento).

Como verificado no caso de estudo anterior, o modelo tende a subestimar as componentes do vento, excepto nas zona em altitude, principalmente na componente v do vento, onde o modelo tende a sobrestimar. Para a intensidade do vento o modelo comporta-se contrariamente às suas componentes, em geral, sobrestimando o vento em todo o pais, com maiores discrepâncias nas zonas em altiude e na zona oeste da costa.

Na tabela 3 estão apresentados os resultados estatísticos para os dois casos de estudo para o prazo de previsão a 36 horas

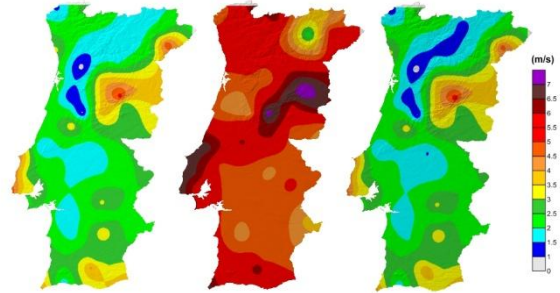
Tabela 3– Resumo dos parâmetros estatísticos de avaliação da qualidade da previsão dos modelos MOS e MM5, para a temperatura e cada componente horizontal do vento.

Casos	Prazo	N°Predictores	Variáveis	Parâmetros Estatísticos				
				BE	r	RMSE	MSE	SC [%]
MM5	36	—	vento	2.62	0.47	4.02	16.18	—
			u	-0.46	0.53	3.69	13.58	
			v	-0.14	0.58	3.66	13.41	
MOS	36	6	vento	-0.025	0.49	2.09	4.38	74
			u	-0.039	0.61	1.78	3.16	77
			v	0.005	0.63	1.63	2.66	80
		17	vento	0.018	0.54	1.98	3.92	76
			u	-0.019	0.67	1.70	2.89	79
			v	0.015	0.67	1.56	2.44	82

Em geral, os resultados mostram uma melhoria do erro médio quadrático na ordem dos 80% ( mais significativa na componente meridional do vento). Contudo, a construção das equações de regressão com 17 predictores mostra uma melhoria na qualidade da previsão, do vento e das respectivas

componentes, ligeiramente mais significativa, traduzida por um aumento de 6% de correlação e menores discrepâncias no erro médio quadrático, quando comparada com a previsão com 6 predictores. Nas figura seguinte estão representados os mapas das médias anuais destes resultados, em comparação com os dados observados.

Vento (36 horas de previsão)  
(Observação, Previsão MM5 e Previsão MOS)



Componente Zonal do Vento (36 horas de previsão)  
(Observação, Previsão MM5 e Previsão MOS)



Componente Zonal do Vento (36 horas de previsão)  
(Observação, Previsão MM5 e Previsão MOS)

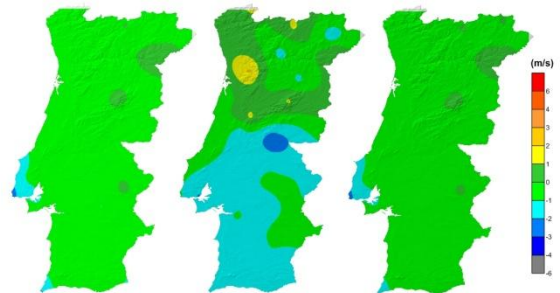


Fig.12.- Resultados, da observação, previsão MM5 e MOS para as 36 horas, com 17 predictores, da média do vento e das suas componentes zonal e meridional.

Nas figuras 13 e 14 estão representados o ciclos médios de 36 horas de duas estações situadas nas regiões (zona oeste e zona em altitude) onde o modelo MM5 tende a sobrestimar com maior discrepâncias as observações.

Desta análise podemos observar que o modelo numérico MM5 não consegue prever com precisão

os efeitos locais. Contudo, o MOS consegue corrigir a intensidade da grandeza simulada bem como as diferenças de fase desta relativamente quando comparada com as observações.

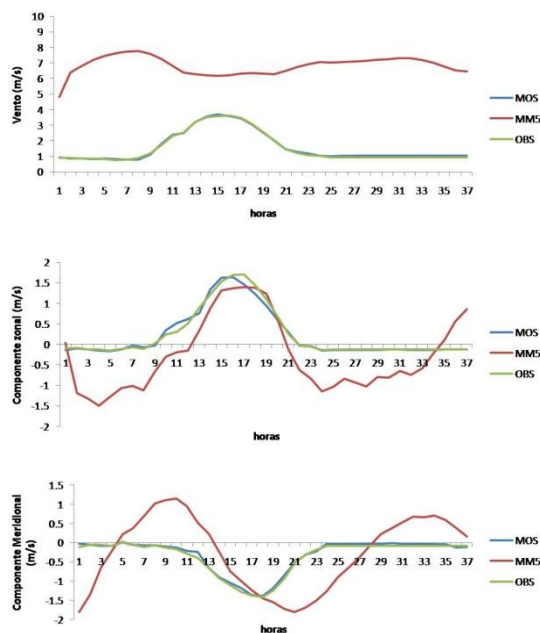


Fig.13.- Ciclo médio de 36 horas do vento e das suas componentes zonal e meridional, observadas e previstas pelos modelos MOS e MM5, na zona do litoral oeste.

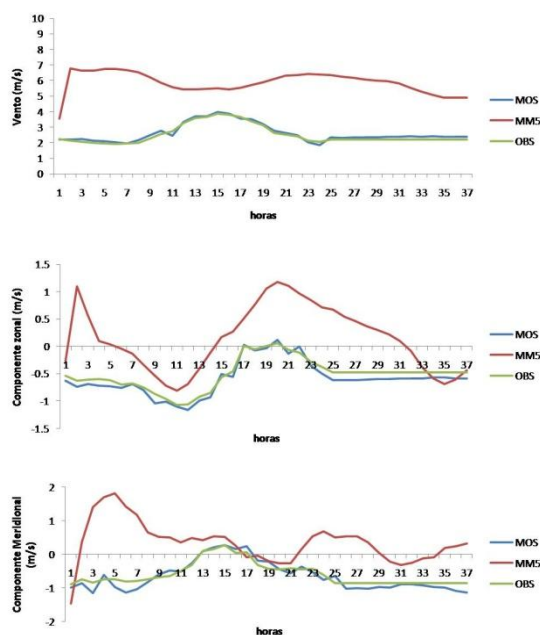


Fig.14.- Ciclo médio de 36 horas do vento e das suas componentes zonal e meridional, observadas e previstas pelos modelos MOS e MM5, numa zona em altitude.

### Conclusões

Os resultados, do estudo inicial, mostram uma melhoria do erro médio quadrático de 80% na previsão das componentes horizontais do vento,

relativamente aos resultados da previsão pontual das 24 horas do modelo numérico MM5. Para as previsões pontuais de 48 e 72 horas, a aplicação do modelo MOS melhora na ordem dos 70% as previsões do MM5.

A aplicação do modelo MOS para o prazo de previsão de 36 horas, aferido em validação cruzada, com espaçamento de 3 e 12h entre os predictores seleccionados para a intensidade do vento e as respectivas componentes, mostra uma melhoria do erro médio quadrático na ordem dos 80%.

Com a utilização de maior número de predictores, na elaboração da recta de regressão, verificasse uma melhoria do erro médio quadrático na ordem dos 2%, para 36 horas de previsão.

Estes resultados motivam a utilização operacional do conceito desta metodologia e o seu desenvolvimento para outras variáveis de previsão.

### Trabalho Futuro

Pretender-se-á continuar o estudo da aplicação desta metodologia com a finalidade de elaborar equações de regressão para fins de previsão operacional e de validação de modelos de mesoscala.

### Agradecimentos

Para a realização deste trabalho agradece-se a colaboração do Professor Pedro Miranda, da Faculdade de Ciências da Universidade de Lisboa, pela cedência dos dados de observação e de previsão MM5 interpolados para a rede IM.

Este trabalho foi realizado no âmbito do Projecto financiado pela FCT - PTDC/CTE-ATM/73607/2006: BRIEF - "Building Regional Ensemble Forecasts".

### Referências

- [1] Storch, H.V., Zwiers, F.W., 1999. Statistical Analysis in Climate Research, University Press, Cambridge, UK.
- [2] Wilks, D.S., 1995. Statistical Methods in Atmospheric Sciences An Introduction, Academic Press, San Diego, California, U.S.A.
- [3] Grell, G. A., J. Dudhia, D. R. Stauffer (1995) A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note, NCAR/TN-398+STR.
- [4] <http://www.weather.ul.pt/>
- [5] Marujo, R., 2009. MOS - Model Output Statistics - Aplicação a Previsões MM5 de curto prazo em Portugal Continental. Dissertação para a obtenção do grau de Mestrado em Ciências Geofísicas - especialização em meteorologia, Lisboa.