A new method to estimate air-quality levels using a synoptic-regression approach. Part I: Present-day O₃ and PM₁₀ analysis

Matthias Demuzere*, Nicole P.M. van Lipzig

Earth and environmental Sciences, Celestijnenlaan 200E, 3001 Heverlee (Leuven), Katholieke Universiteit Leuven, Belgium

ABSTRACT

In order to make projections for future air-quality levels, a robust methodology is needed that succeeds in reconstructing present-day air-quality levels. At present, climate projections for meteorological variables are available from Atmospheric-Ocean Coupled Global Climate Models (AOGCMs) but the temporal and spatial resolution is insufficient for air-quality assessment. Therefore, a variety of methods are tested in this paper in their ability to hindcast maximum 8 hourly levels of O₃ and daily mean PM₁₀ from observed meteorological data. The methods are based on a multiple linear regression technique combined with the automated Lamb weather classification. Moreover, we studied whether the above-mentioned multiple regression analysis still holds when driven by operational ECMWF (European Center for Medium-Range Weather Forecast) meteorological data. The main results show that a weather type classification prior to the regression analysis is superior to a simple linear regression approach. In contrast to PM₁₀ downscaling, seasonal characteristics should be taken into account during the downscaling of O₃ time series. Apart from a lower explained variance due to intrinsic limitations of the regression approach itself, a lower variability of the meteorological predictors (resolution effect) and model deficiencies, this synoptic-regression-based tool is generally able to reproduce the relevant statistical properties of the observed O₃ distributions important in terms of European air quality Directives and air quality mitigation strategies. For PM₁₀, the situation is different as the approach using only meteorology data was found to be insufficient to explain the observed PM₁₀ variability using the meteorological variables considered in this study.

1. Introduction

Since high concentrations of O₃ and PM₁₀ affect the public health, much attention is paid to the improvement of the accuracy of short-term deterministic and statistical prediction models and the development of robust long-term air-quality prediction systems. Complex models, including a full description of atmospheric chemistry and meteorological processes, are often used with respect to the former (Giorgi and Meleux, 2007). Although these techniques are shown to be powerful for short-term predictions, the complex climate–air-quality modelling systems, together with their computational/technical characteristics, are at present less useful for long-term (decadal) predictions (Giorgi and Meleux, 2007).

Therefore, statistical downscaling methods were developed to determine predictive relationships between air pollution concentrations and individual meteorological parameters. This was done for different air-quality variables and different locations. Different methods can be distinguished namely multiple linear regression (MLR) analysis (Barrero et al., 2006), nonlinear multiple regressions (Cobourn, 2007), artificial neural networks (ANN) (e.g. Hooyberghs et al., 2005), and generalized additive models and fuzzy-logic-based models (Cobourn et al., 2000). Among these approaches, regression methods are well documented because of their ease of implementation and their low computation requirements (Wilby and Wigley, 1997; Wilby et al., 2004). The technique first detects present-day relationships between local meteorological variables (or "predictors") and e.g. air-quality variables (or "predictands"). These relations based on (non) linear multiple regressions have been described in literature numerously, as for e.g. in Chaloulakou et al. (2003) and Ainslie and Steyn (2007). Nevertheless, the use of another circulation patterns as a downscaling tool is widespread, although less common in air-quality research. In that respect, this technique is adopted by e.g. Comrie and Yarnal (1992) and Davies et al. (1992) to explain observed ozone variability at measuring sites in the Europe, US and Canada. Some authors

* Corresponding author. Tel.: +32 16326424; fax: +32 16326400.
E-mail address: matthias.demuzere@ees.kuleuven.be (M. Demuzere).

have tried to combine the above-mentioned two downscaling techniques. A stratification based on the circulation pattern is adopted to introduce nonlinearity into the model (Huth et al., 2008a,b) under the assumption that the relations between large-scale predictors and predictands may vary depending on the type of the synoptic pattern. This technique is applied for downscaling surface meteorological variables by e.g. Cavazos (1999). Nevertheless, this approach has not been adopted in terms of air quality assessment yet. To the author's best knowledge, there has been only one study so far that used a within-synoptic-type air pollution model to study future air pollution levels for a variety of pollutants (Cheng et al., 2007a,b).

The aim of our study is to test a simple linear regression method together with a stratification of the dataset by its daily synoptic patterns in their ability to hindcast levels of O_3 and PM_{10}. To achieve this goal, regression results from previous research based on station measurements from Cabauw (Demuzere et al., 2009) are used to reconstruct the observed maximum eight hourly mean O_3 [$\mu g\ m^{-3}$] (hereafter referred to as $m8O_3$) and daily mean PM_{10} [$\mu g\ m^{-3}$] levels for the period 2001–2004 (calibration) and 2005–2006 (evaluation). Furthermore, the Lamb weather type classification is used as a synoptic circulation-typing tool to enter nonlinearity into the downscaling model in order to construct a robust method for the improved projections of air-quality levels. These methods are calibrated on observed air quality data from the Cabauw measurement tower and the measurement station of Zegveld-Oude Meije (The Netherlands) for the period 2001–2004.

According to the Intergovernmental of Climate Change, the magnitude of the effect of climate change on O_3 is still uncertain (IPCC, 2001). Hence, before the above-mentioned downscaling approach is used to bridge the gap between what is produced by AOGCMs (Atmospheric–Ocean Coupled Global Climate Models) and what is needed in climate impact research (Part II of this analysis), the observed local relations between meteorological and air-quality variables need to be tested on the larger scale. After all, AOGCM-output is at present only available typically scales of $300 \times 300\ km^2$. Therefore, the validity of these station-based synoptic-regression configurations as an air quality downscaling tool is tested using low-resolution operational ECMWF (European Center for Medium-Range Weather Forecast) data for the period 2005–2006. In this way, a regression-based model forced with meteorological ECMWF data is used as a prototype for a modelling system in which AOGCM-output is downscaled for the purpose of obtaining projections for future air-quality levels.

2. Data

High temporal resolution meteorological data for the period 2001–2006 from the rural measurement station of Cabauw (The Netherlands), partly operated by the KNMI (Royal Dutch Meteorological Institute were used. Ten-minute measurements are averaged to daily values. More details on the measurement site characteristics and quality control are provided in Demuzere et al. (2009). The calibration of the multiple linear regression equation is based on local observations of 2 m air and dew point temperature (T and Td), daily maximum and minimum temperature (T_{max} and T_{min}), wind speed and direction (F010 and D010), total cloud cover (CC), total precipitation (T_{prec}) and shortwave downward radiation (SDW) for the period 2001–2006. Relative humidity (RH) is calculated from T and Td using the Magnus-Tetens approximation. Similar meteorological information is obtained from the operational ECMWF model on an N400 Gaussian grid resolution for the period 2005–2006. As this methodology is developed for providing future air-quality projections based on global climate model information, the ECMWF data is aggregated to the spectral T63

resolution ($1.85° \times 1.85°$), which is identical to the resolution of e.g. the SRESA1B IPCC 4AR ECHAM5-MPI/OM experiment (Marsland et al., 2003; Roeckner et al., 2003). As Cabauw is situated in the center of 4 surrounding grid points, the average of the neighbouring 4 ECMWF grid points is used in the further analysis, as Gachon and Dibike (2007) have shown that for a range of meteorological variables there is virtually no difference between a single grid point use and neighbouring grid point averages.

The air-quality data for ozone is taken from Cabauw, for the period 2001–2006, while PM_{10} is taken from the neighbouring station Zegveld-Oude Meije. The Directive 1999/30/EC and following up Directive 2008/50/EC of the European Parliament (EU, 2008) describe a threshold concentration of $120\ \mu g\ m^{-3}$ and $50\ \mu g\ m^{-3}$ for $m8O_3$ and daily mean PM_{10} concentration respectively for Europe. These threshold values are important for possible regional air-quality mitigation strategies and therefore, the remaining of this study focuses on these thresholds as the framework to validate the downscaling methodologies provided in Section 3.

3. Methods

Multiple linear regression models have widely been adopted to reconstruct observed time series of O_3 and PM_{10} for various heterogeneous regions based on measurements (Barrero et al., 2006). In addition, Huth et al. (2008a,b) points out the power of a pointwise linear regression method (using grid point values instead of principal components as predictors) in comparison with nonlinear methods such as neural networks. And although Huth et al. (2008a,b) do not find clear evidence in the improvement of downscaling temperatures when stratifying the dataset by classification patterns prior to the regression analysis, Cheng et al. (2007a) have shown this approach to be promising in terms of air pollution variables. Therefore, in order to clarify the potential of each method in air-quality applications, three regression-based approaches are examined and compared: (1) a multiple linear regression model as developed in Demuzere et al. (2009), hereafter named MLR, (2) a multiple regression model with the Lamb weather types (Jones et al., 1993) included as predictors (MLR$_{LWT}$) and (3) a stratification of the data by the Lamb weather type technique prior to the multiple linear regression (LWT$_{MLR}$). As seasonality could be an important factor in terms of $m8O_3$ and PM_{10} reconstruction (Tarasova et al., 2007), model approaches (2) and (3) are also run in a seasonal mode, whereby the regression analysis is done for each season separately. These techniques are hereafter referred to as MLR$_{SEAS}$ and LWT$_{SEAS+MLR}$ for (2) and (3) respectively. The latter results in a set of N_{LWT} times N_{SEAS} regression equations and describes the seasonal within-weather-type air pollution characteristics, where N_{LWT} is the number of weather type classes and N_{SEAS} is the number of seasons (44 equations in total). As an example, Table 1 enlists for each season the number of occurrence

Table 1
The number of predictor occurrences in the synoptic-regression approach over all circulation types and for each season (DJF – winter, MAM – spring, JJA – summer, SON – autumn).

	DJF	MAM	JJA	SON
Sea-Level Pressure (P0)	7	1	2	1
Total Precipitation (T_{prec})	2	1	3	2
Shortwave Downward Radiation (SWD)	7	7	6	4
Maximum Temperature (T_{max})	3	3	8	4
Minimum Temperature (T_{min})	1	1	0	2
Relative Humidity (RH)	3	6	6	6
10m Wind Speed (F010)	8	3	2	3
Wind Direction (D010)	5	1	4	4
Total Cloud Cover (CC)	4	2	2	2

of each predictor in the 44 linear equations. Over all seasons, shortwave downward radiation, relative humidity and maximum temperature are the most frequently occurring predictors, with the largest individual frequency for the latter in JJA (72.7%). In winter, wind speed is equally important as T_{max} in summer, while in the transition seasons MAM (spring) and SON (autumn), the frequencies of occurrence are similar for each predictor variable.

Moreover, it is important to underline that air-quality variables are not taken into account as predictor variables. As there is no information on future air-quality data as input variables from AOGCMs, they cannot be included in the regression analysis as predictors. For future use, this approach has two important assumptions: 1) when applying these techniques for future air-quality projections, the impact of meteorological changes on future O_3 and PM_{10} levels is isolated, and possible changes in future (precursor) emissions are not assessed and 2) applying this method on future AOGCM scenarios assumes that the present-day relations between meteorology and air-quality variables stay constant through time.

The Lamb weather types (WTs) are developed using operational ECMWF sea-level pressure (SLP) data for the calibration period 2001–2004 and for the validation period 2005–2006. For a given day, the WTs describe the location of the high and low-pressure centers that determine the direction of the geostrophic flow. A grid with 16 points is assigned over the larger Western and Central Europe, with a central point over the Netherlands, in $52.5°N$ and $5°E$. We computed a set of simple atmospheric circulation indices using 12 h sea-level pressure in these 16 grid points, namely the direction and vorticity of geostrophic flow: southerly flow (SF), westerly flow (WF), total flow (F), southerly shear vorticity (ZS), westerly shear vorticity (ZW) and total shear vorticity (Z). A small number of empirical rules devised previously (Jones et al., 1993) are then used to classify each day as one of the $10 + 1$ undetermined circulation types as developed in Demuzere et al. (2008, 2009). The composite maps of SLP for each Lamb weather type derived from ECMWF fields, averaged over the period 2001–2006 are depicted in Fig. 1.

A wide range of possible criteria can be used to evaluate statistical downscaling methods. First, the quality and reliability of the downscaled values are examined using several statistical indices proposed by Willmott (1981) and Willmott and Matsuura (2005). These indices include the mean, the standard deviation (σ) and variance (σ^2) and the mean absolute error (MAE), which averages the absolute error magnitudes. Furthermore, we used the root mean square error (RMSE), measuring the total deviation of downscaled values from observed values and the explained variance (R^2). In addition, higher-order moments of the distribution are essential for air-quality modelling, where adverse health effects are more related exceedences of thresholds than to mean air-quality levels. Therefore, as suggested by Huth et al. (2008a,b), the degree of asymmetry and peakedness are evaluated in terms of the third and fourth moments, viz. skewness and kurtosis respectively.

Furthermore, Murphy (1988) suggested that the skill of any given model is a measure of the relative accuracy of a model with respect to a standard reference model. Hence, the skill of any model should be interpreted as the percentage improvement over a reference model. One of the most common used reference models in forecasts is persistence i.e., the previous value in a sequence of observations (Wilks, 1995). Therefore, the skill scores based on the mean square error (MSE) will be used in this paper with persistence (MSE_{pers}) as a reference:

$$SS_p(MSE) = \frac{MSE - MSE_{pers}}{0 - MSE_{pers}} \times 100\%$$

with MSE = 0 corresponding to the accuracy level that would be achieved by a perfect model (Wilks, 1995). Furthermore, the student t-test and the F-statistics are used to test whether observed and downscaled time series have significant different means and variances respectively, on a 95% confidence level.

4. Results and discussion

First, the various multiple linear regression equations described in Section 3 are calibrated for the period 2001–2004, using observational data (Section 4.1). Secondly, it is tested which of the methods is most suitable for hindcasting $m8O_3$ and daily mean PM_{10} concentrations for the independent evaluation time period 2005–2006, using observed meteorological data as input to the regression-based model. Thirdly, it is tested whether the MLR equations derived from observed meteorological data still hold when using gridded low-resolution meteorological data as input to the regression-based model. For this purpose operational ECMWF meteorological data for the 2005–2006 are used (Section 4.2). Finally, Section 4.3 describes the within-type variability of the meteorogical predictors provided by Cabauw observations and ECMWF forecast data.

4.1. Comparison and evaluation of various downscaling tools

In order to justify the implementation of a circulation approach prior to the linear regression technique, it is important to know whether the introduction of this stratification leads to an improvement in performance of the downscaling procedure. Therefore, the comparisons are carried out not only in terms of correlation coefficients, but also for temporal correlations (persistence) and higher-order statistical moments (skewness and kurtosis). These performance statistics for the calibration and evaluation period are summarized in Tables 2–4 respectively. In general, the combination of synoptic classification using Lamb weather types prior to a seasonal linear multiple regression analysis ($LWT_{SEAS+MLR}$) performs best for both $m8O_3$ and PM_{10}. This is shown in terms of explained variance and SS_p for both $m8O_3$ and PM_{10} (80/75% and 60/66% respectively). Furthermore, MAE and RMSE show a similar tendency, with the lowest values for $m8O_3$ of 8.3 and 10.9 $\mu g\ m^{-3}$ and for PM_{10} of 5.8 and 8.6 $\mu g\ m^{-3}$ compared to the other approaches. Differences between observed and modelled means are insignificant for all model configurations based on the t-test. In terms of variance, the F-test shows that all models are unable to reproduce the observed variance, except for $m8O_3$ using MLR_{SEAS}, LWT_{MLR} and $LWT_{SEAS+MLR}$. As all statistics for PM_{10} are lower than for $m8O_3$, these results suggest that $m8O_3$ has more predictability based on meteorology solely compared to PM_{10}.

To further illustrate this, scatter plots of modelled versus observed $m8O_3$ (upper panels) and PM_{10} (lower panels) concentrations following the pure MLR (left panels) and synoptic-regression (right panels) approach are shown (Fig. 2). For ozone, the best results are obtained for $LWT_{SEAS+MLR}$, with the regression line closer to the 1:1 (perfect model) line. For PM_{10}, there is a large difference between the various models, with again the best performance for the $LWT_{SEAS+MLR}$ model.

Furthermore, Tables 2–4 show the models performance for hindcasting $m8O_3$ and daily mean PM_{10} levels for the independent evaluation time period 2005–2006, using observed meteorological variables. For $m8O_3$, all models perform similar in terms of R^2 and SS_p, with the best results obtained for R^2 for $LWT_{SEAS+MLR}$ (72%) and for SS_p with MLR_{SEAS} (59%). Both model approaches are also able to reproduce the observed variances, whereas all other model configurations have variances significant different on the 95% confidence interval compared to the observed explained variance. In terms of skewness and kurtosis, the best results are obtained
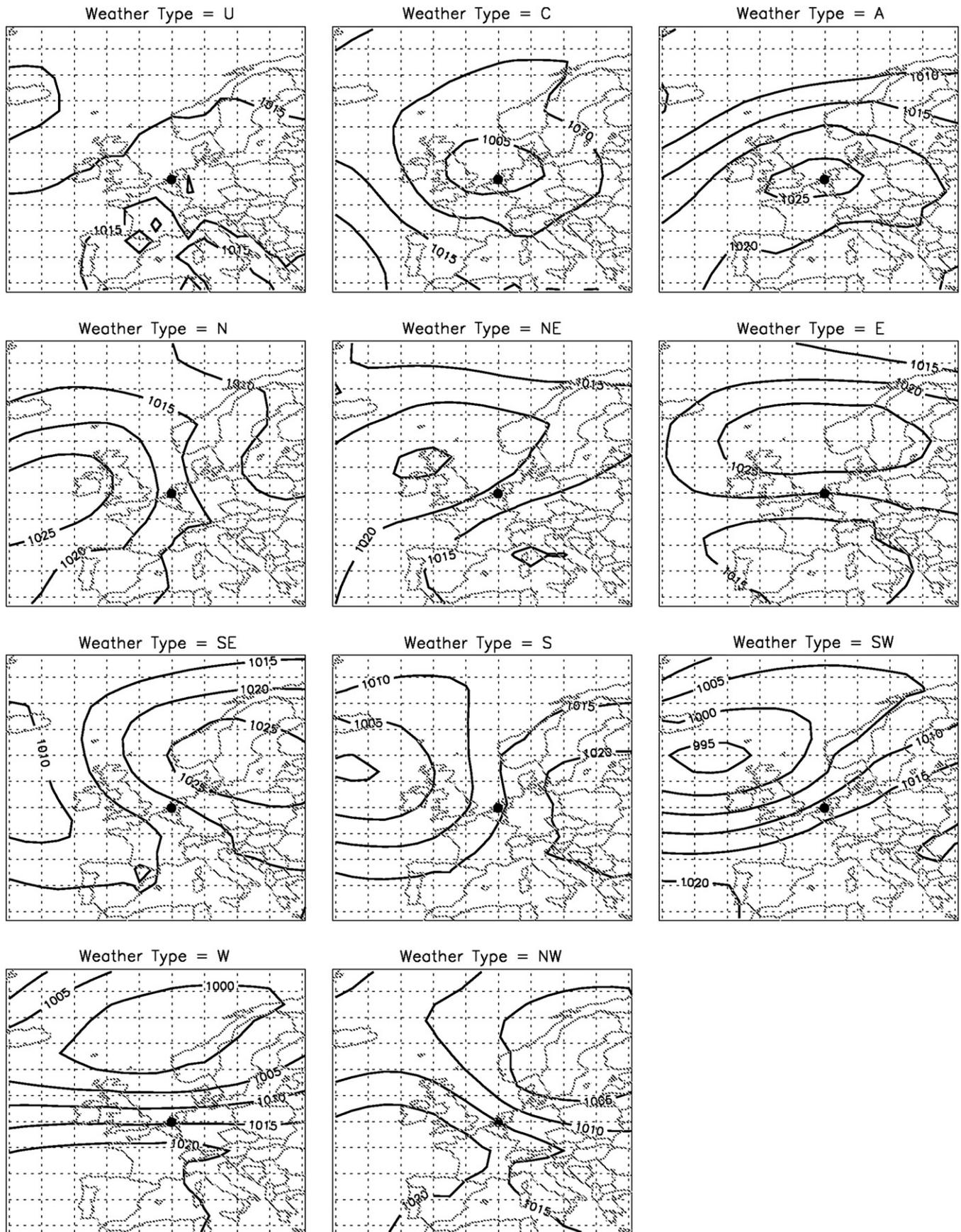
Fig. 1. Composite maps of sea-level pressure for each Lamb weather type derived from ECMWF SLP fields, averaged over the whole period 2001–2006. The Cabauw measurement site is marked with a black dot.

Table 2
Statistics describing the performance of the calibration time series (2001–2004) and evaluation series (2005–2006) for ozone and PM_{10} ($\mu g\ m^{-3}$). The measures include the mean ($\bar{x}$), standard deviation (σ), variance (σ^2), mean absolute error (MAE), root mean square error (RMSE), explained variance (R^2 in %) and skill score against persistence (SSp in %). Values significant different on the 95% confidence interval based on t-test and chi-square test for mean and variance respectively are denoted in bold.

	Species	Model		$\bar{x}$	σ	σ^2	MAE	RMSE	R^2	SSp
Calibration 2001–2004	O_3		*Reference*	*57.0*	*29.8*	*887.2*				
		MLR		54.9	25.9	**671.4**	14.3	19.0	61	38
		MLR_SEAS		55.1	27.5	758.7	10.4	13.2	65	69
		MLR_LWT		56.6	26.0	**676.3**	12.6	15.9	71	56
		LWT_MLR		55.2	27.4	751.5	10.2	13.9	78	67
		LWT_SEAS+MLR		56.5	28.0	783.1	8.3	10.9	80	75
	PM_{10}		*Reference*	*27.2*	*15.6*	*244.6*				
		MLR		27.7	9.1	**83.4**	11.1	15.1	12	22
		MLR_SEAS		27.1	10.5	**110.5**	8.6	11.8	41	50
		MLR_LWT		27.0	9.9	**97.8**	8.6	11.8	40	52
		LWT_MLR		27.0	12.5	**155.3**	3.4	10.8	52	62
		LWT_SEAS+MLR		27.1	13.4	**180.6**	5.8	8.6	60	66
	O_3		*Reference*	*57.59*	*30.82*	*955.90*				
		MLR	*Observations*	56.60	25.28	**639.29**	13.70	17.70	68	50
			ECMWF	58.05	21.52	**463.13**	16.98	22.42	60	37
		MLR_SEAS	*Observations*	57.13	27.87	776.57	11.77	15.11	58	59
			ECMWF	58.24	22.03	**485.32**	13.97	17.92	44	43
		MLR_LWT	*Observations*	56.48	26.55	**704.85**	13.46	17.65	68	51
			ECMWF	**63.35**	26.62	**708.37**	16.00	20.22	61	34
		LWT_MLR	*Observations*	57.34	26.81	**718.59**	13.13	17.35	69	52
			ECMWF	**59.42**	27.98	782.89	19.68	25.76	61	31
		LWT_SEAS+MLR	*Observations*	57.87	29.37	862.85	12.76	16.71	72	56
			ECMWF	57.6	27.14	**736.58**	16.24	20.42	63	33
Evaluation 2005–2006	PM_{10}		*Reference*	*23.13*	*12.05*	*145.20*				
		MLR	*Observations*	**26.64**	8.17	**66.71**	7.16	10.29	32	35
			ECMWF	**21.45**	8.31	**69.04**	8.59	10.72	28	43
		MLR_SEAS	*Observations*	**25.56**	10.20	**104.09**	8.39	10.62	34	30
			ECMWF	**20.72**	9.76	**95.27**	8.35	11.20	27	27
		MLR_LWT	*Observations*	**25.18**	10.54	**111.17**	8.47	10.73	31	35
			ECMWF	**19.16**	10.80	**116.57**	8.45	10.73	32	31
		LWT_MLR	*Observations*	**26.16**	12.66	160.15	8.32	10.54	34	34
			ECMWF	23.03	13.78	189.88	8.40	10.62	32	28
		LWT_SEAS+MLR	*Observations*	**26.75**	15.98	**255.30**	10.55	15.22	34	33
			ECMWF	**24.34**	15.74	247.75	10.17	15.22	32	28

with the LWT_SEAS+MLR model, with values of 0.91 and 3.27 respectively, compared to an observed skewness and kurtosis of 1 and 2.28 respectively (Tables 3 and 4). For PM_{10} in general, lower values of the statistical indices are obtained compared to the results for m8O3. Again here, models perform similar in terms of R^2 and SS_p for all methods, with the best results obtained with MLR_SEAS, MLR_LWT and LWT_SEAS+MLR for the former (34%) and for MLR and MLR_LWT for the latter (35%). The large observed skewness and kurtosis values are best reproduced by LWT_MLR with respective values of 1.02 and 4.29 compared to an observed skewness and

kurtosis of 1.73 and 4.78 respectively. These results are also depicted in Fig. 3 (upper panels), showing the best representation of observed m8O3 and daily mean PM_{10} levels based on observed meteorological variables for the independent period 2005–2006 using LWT_SEAS+MLR and LWT_MLR respectively.

4.2. Evaluation of various downscaling tools for low-resolution data

In order to know whether regression-based models developed from measured predictors are still valid using lower resolution data,

Table 3
Skewness coefficients for O_3 (left) and PM_{10} (right) for the various methods and the reference and observed and ECMWF downscaled datasets, for the period 2005–2006.

	O_3			PM_{10}		
	Reference	Obs	ECMWF	Reference	Obs	ECMWF
	1.00			*1.73*		
MLR		0.26	−0.002		−0.085	0.043
MLR_SEAS		0.78	0.42		−0.19	−0.11
MLR_LWT		0.28	0.15		−0.27	−0.056
LWT_MLR		0.62	0.31		1.02	0.70
LWT_SEAS+MLR		0.91	0.80		3.08	0.82

Table 4
Kurtosis coefficients for O_3 (left) and PM_{10} (right) for the various methods and the reference and observed and ECMWF downscaled datasets, for the period 2005–2006.

	O_3			PM_{10}		
	Reference	Obs	ECMWF	Reference	Obs	ECMWF
	2.28			*4.78*		
MLR		−0.40	−0.71		−0.083	−0.50
MLR_SEAS		0.86	0.43		0.12	−0.20
MLR_LWT		−0.27	−0.63		−0.26	−0.49
LWT_MLR		0.51	0.028		4.29	6.54
LWT_SEAS+MLR		3.27	2.31		6.14	6.92
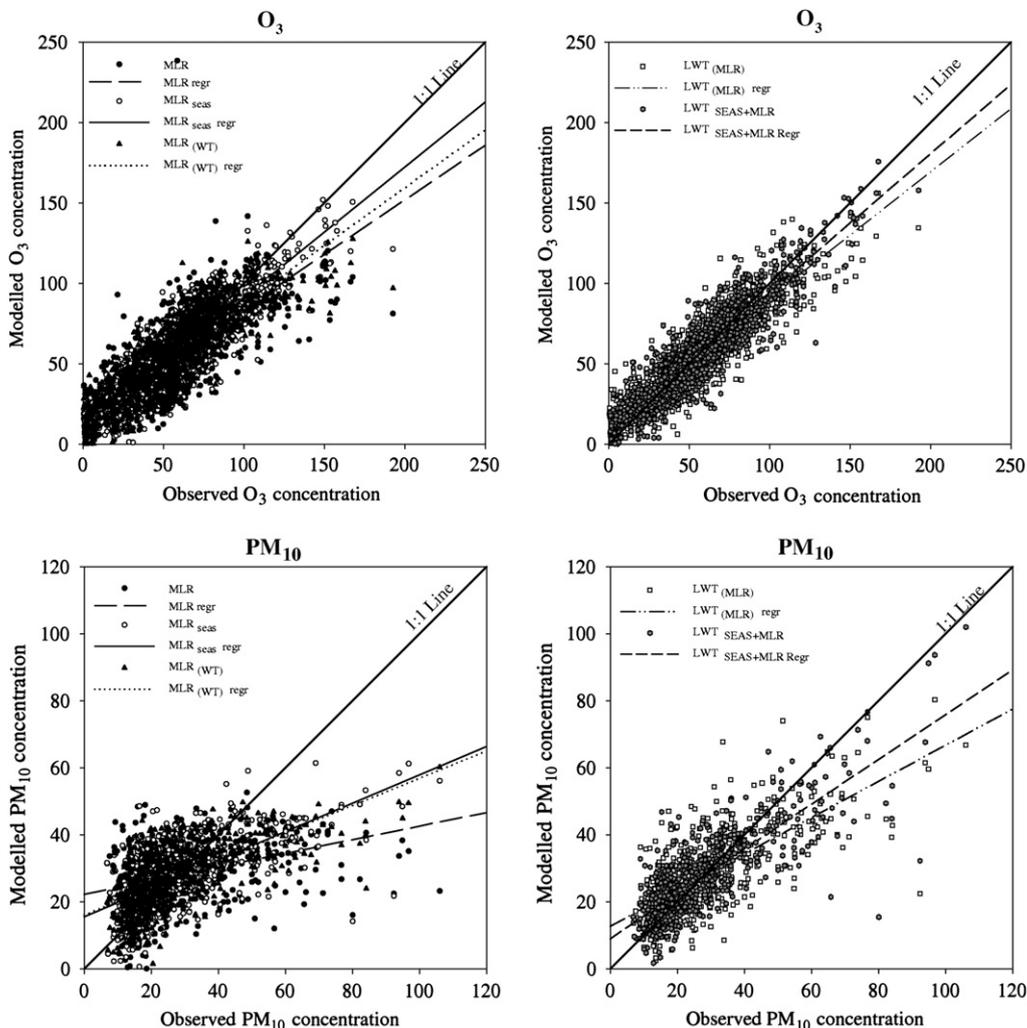
Fig. 2. Scatter plot of modelled versus observed eight hourly maximum O_3 (upper panels) and daily mean PM_{10} (lower panels) concentrations [µg m^{-3}] for the calibration period 2001–2004.

such as operational ECMWF data, this approach is tested on the latter for the independent evaluation period 2005–2006 (Fig. 3 – lower panels). In general the lower variability of the low-resolution meteorological ECMWF predictors results in a smaller explained variance of operational ECMWF downscaled $m8O_3$ and PM_{10} time series compared to the observed downscaled series (Table 2). The explained variance for $m8O_3$ is, in accordance to the calibration period, the highest for the LWT$_{SEAS+MLR}$ method for both observed and ECMWF downscaled series (72 and 63%). In terms of SS_p, the pure seasonal regression model MLR$_{SEAS}$ is slightly better than the LWT$_{SEAS+MLR}$ method, with a performance of 59/43% and 56/33% (Observations/ECMWF) respectively against persistence (Table 2).

In order to compare the distribution of the observed and modelled $m8O_3$ and daily mean PM_{10} throughout the evaluation period, the quantile distribution of the modelled air-quality data is plotted against the quantile distribution of the observed air-quality data in qq-plots (Fig. 3). When using the ECMWF data as predictors, the LWT$_{MLR}$ and LWT$_{SEAS+MLR}$ approaches perform similar for the higher percentiles, although the former has a larger overestimation of modelled $m8O_3$ concentrations around 100 µg m^{-3} (Fig. 3 – lower panel). Moreover, the degree of asymmetry and peakedness of the statistical distributions are also evaluated in terms of the third and fourth order statistical moments, i.e. skewness and kurtosis (Tables 3 and 4). Following the supposition that the number of independent

realizations in the time series is 70 (approximately every 10th day as an independent realization), than the skewness test for normality (Thode, 2002) indicates that the hypothesis of zero skewness is rejected at the 95% significance level if the skewness coefficient exceeds a value of 0.459 (in the absolute sense). The reference O_3 distribution shows a positive skewness of 1.00 (significant skewed) and a kurtosis of 2.28. When using observed meteorological data as predictors, the methods applying stratification of the $m8O_3$ distribution by the Lamb classification method prior to the regression analysis (LWT$_{SEAS+MLR}$) are able to catch the third and fourth order statistical measures much better than methods without such stratification. Similarly, the same conclusions can be drawn when using the ECMWF data as predictors, although the distribution is slightly less positive skewed in that case (Table 4).

For PM_{10}, a regression-based approach using only meteorology data is insufficient to explain a great deal of the observed PM_{10} variability (Table 2). The results of the regression-based models are similar in terms of R^2, whereby the best approximation of the observed time series is obtained by both LWT$_{MLR}$ and LWT$_{SEAS+MLR}$ (34 and 34% respectively). Contrarily, better results for SS_p are obtained for the standard regression model MLR, with a 35% and 43% improvement for both observed and ECMWF downscaled PM_{10} series compared to the persistence model. For all models except LWT$_{MLR}$ the variance (σ^2) differs significantly
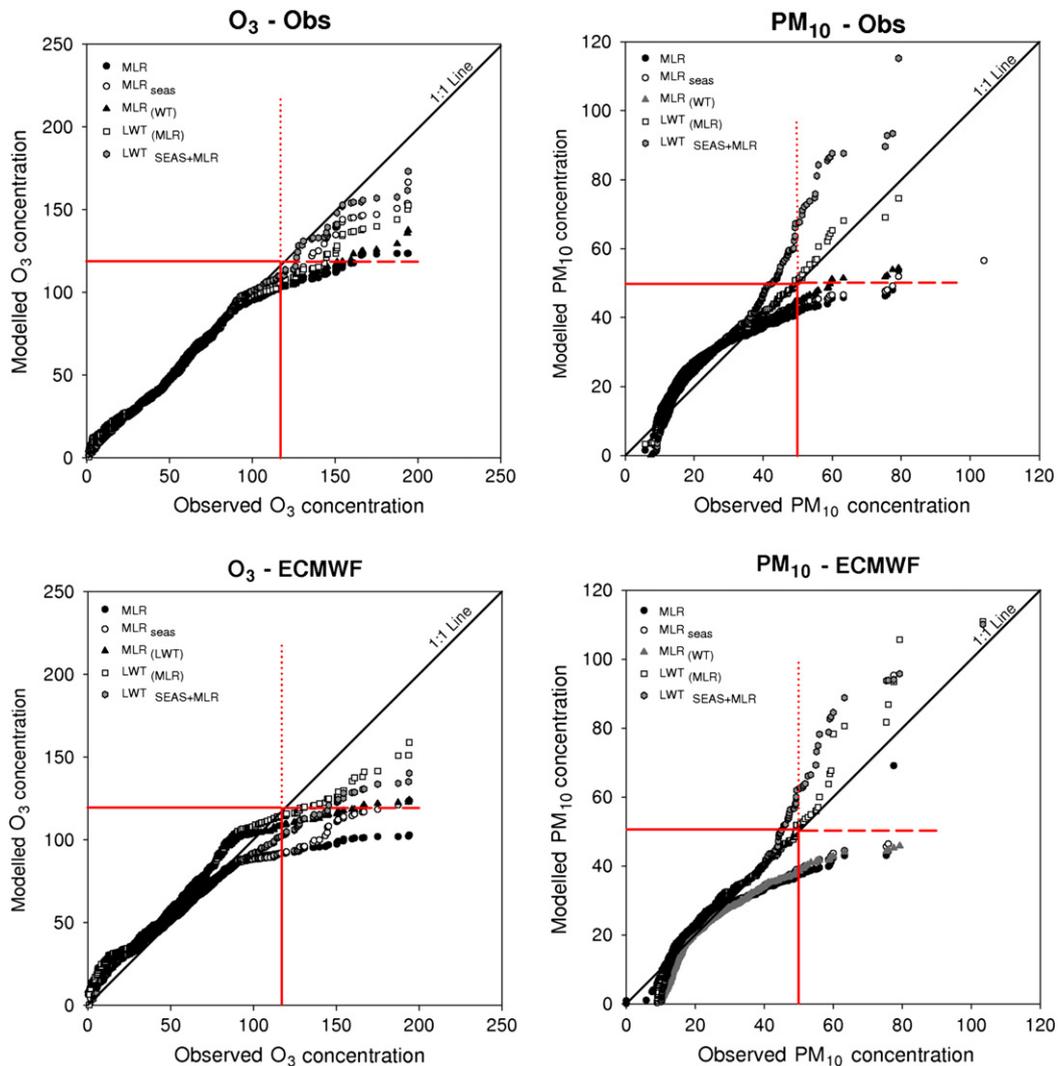
Fig. 3. Quantile–quantile plots of modelled versus observed eight hourly maximum O_3 and daily mean PM_{10} concentrations [$\mu g\ m^{-3}$] for the evaluation period 2005–2006, using observed (upper panels) and operational ECMWF (lower panels) data as predictors for the regression-based models. The 1:1 line presents the perfect model hindcast and the red lines (dotted – observation, dashed – modelled) show the threshold of 120 and 50 $\mu g\ m^{-3}$ for O_3 and PM_{10} respectively.

from the observed variance in PM_{10}. The reference PM_{10} distribution is highly positively skewed (significant on the 95% level), which is well reproduced by the LWT_{MLR} approach and overestimated by $LWT_{SEAS+MLR}$ when observed meteorological data are used as predictors (Tables 3 and 4). Moreover, the LWT_{MLR} model more realistically generates the strong reference peakedness of 4.775 compared to the $LWT_{SEAS+MLR}$ approach (Table 4). For the ECMWF downscaled PM_{10} time series, the distribution is in general more normal compared to the reference distribution, although the kurtosis is overestimated for the LWT_{MLR} and $LWT_{SEAS+MLR}$ approach (Tables 3 and 4).

In general, the above-analysis shows that for $m8O_3$, seasonality is an important factor, which confirms the results of Tarasova et al. (2007). Next to its photochemical production and the seasonal variation of anthropogenic and biogenic emissions, a season-dependant impact of meteorological variables (e.g. wind speed and total precipitation) on ozone should be taken into account in the downscaling process. For PM_{10}, this effect is less distinct, which is also clear from a less pronounced seasonal cycle in observed PM_{10} concentrations (Flemming et al., 2005; Demuzere et al., 2009). Moreover, the results in Table 2 show that a regression-based

approach using only meteorology data is able to explain a great deal of the observed $m8O_3$ variability, which is not the case for PM_{10}. Possibly, a part of the low explained variability is due to the absence of the boundary layer height as a predictor variable. Previous research (Hooyberghs et al., 2005) has shown that by far the boundary layer height is the most important parameter in their neural network approach. Unfortunately, boundary layer height measurements are not available at present for Cabauw.

Furthermore, the results in Tables 2–4 show an additional value of stratifying all days according to their circulation characteristics prior to the regression analysis. Although Huth et al. (2008a,b) states that the effect of introducing non-normality by mixing several normal distributed meteorological parameters using a classification method for temperature downscaling is rather weak and insufficient to produce significant deviations from normality, our result show that this is not the case for air-quality distributions. Hereby, large skewness and kurtosis values can only be reproduced by stratifying the dataset by developing a Lamb weather-type-dependant regression model as described in Section 3. Moreover, the skill score, which tests the temporal structure of the downscaled time series against a 1-day lag persistence model, shows

a better performance for the LWT methods than compared to the more simple linear regression model approaches (Table 2). Therefore we opt to use the $LWT_{MLR+SEAS}$ and LWT_{MLR} for $m8O_3$ and PM_{10} respectively in the remaining of this study.

Finally, the hypothesis that an observation-based regression still holds when applied on low-resolution gridded (ECMWF) data is validated and shown in Tables 2–4. Although part of the explained variance is lost due to a lower variability of the meteorological
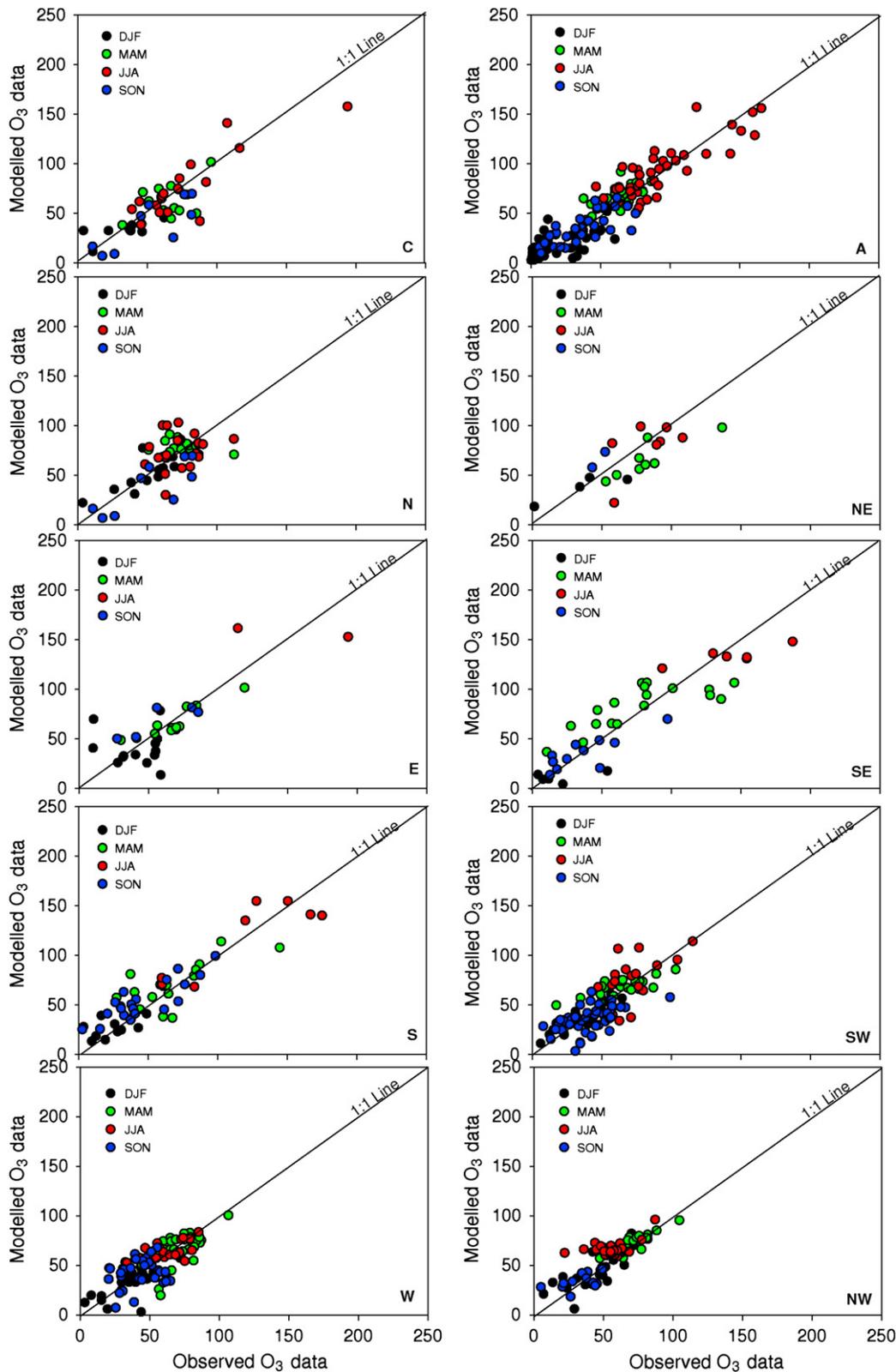
Fig. 4. Scatter plots showing observed versus modelled O_3 concentrations using observed meteorological data as input to the $LWT_{SEAS+MLR}$ model for each Lamb weather type and for each season separately. (For interpretation of the colour plots and legend, the reader is referred to the web version of this article.)

predictors (Fig. 3), this regression-circulation pattern approach is able to detect the high observed m8O$_3$ and PM$_{10}$ concentrations that are important in terms of European air quality Directives and air quality mitigation strategies. This will be discussed in more detail the following Section 4.3.

4.3. Meteorological predictor variability

In the previous paragraph, regression-based models were used to hindcast air-quality levels by calculating m8O$_3$ and PM$_{10}$ concentrations from meteorological conditions on a larger scale as
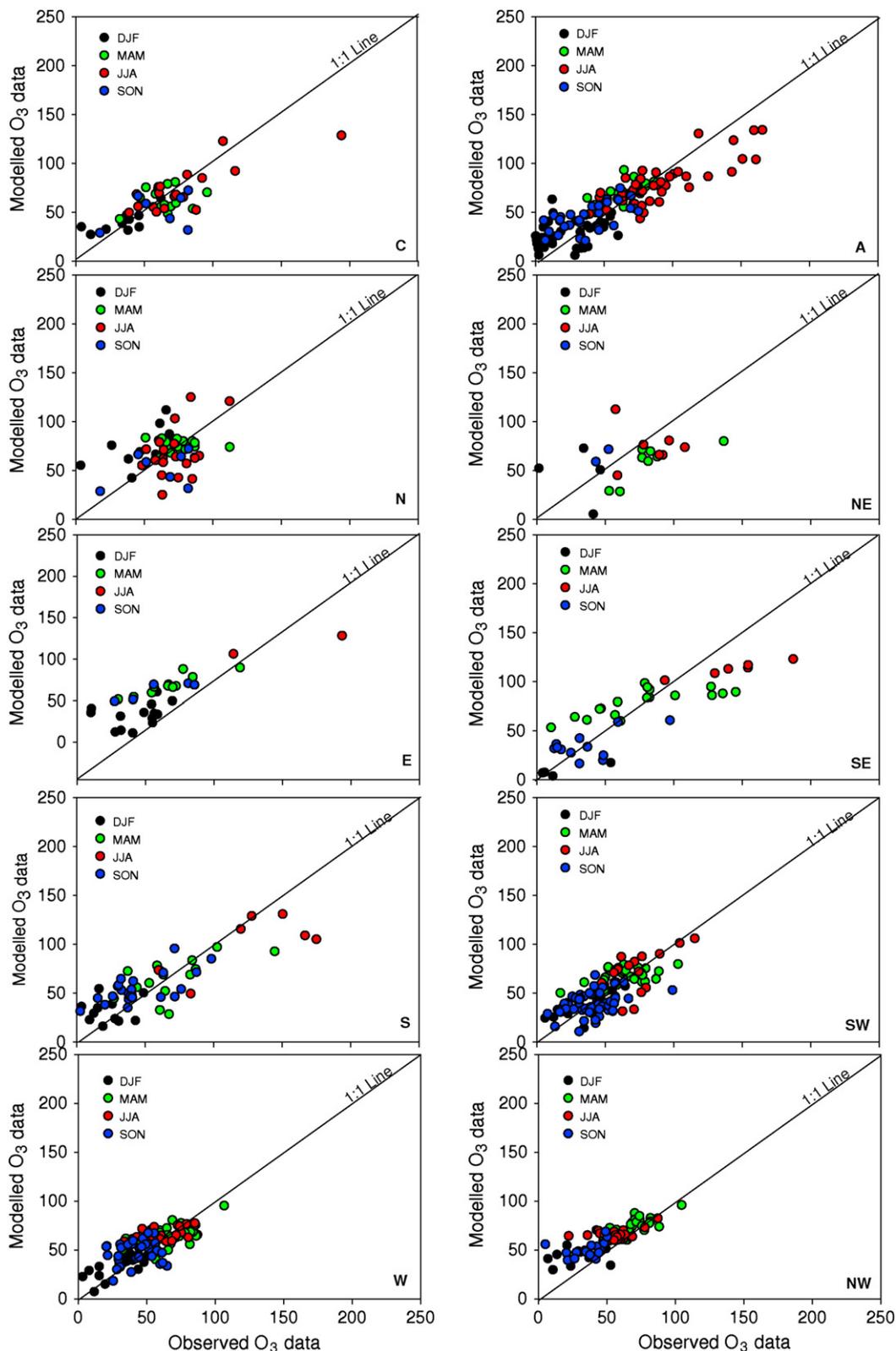
Fig. 5. Same as Fig. 4, but now using ECMWF meteorological data as input to the LWT$_{SEAS+MLR}$ model. (For interpretation of the colour plots and legend, the reader is referred to the web version of this article.)

available from ECMWF. However, when the ECMWF output differs from the observed conditions, this will negatively influence the model performance. Therefore, it is important to provide more insight in the effect of forecasted meteorological ECMWF variable deficiencies on the performance of the regression-based models for each weather type separately. As the regression-based method profits from introducing nonlinearity by the objective Lamb classification method prior to the regression analysis, modelled versus
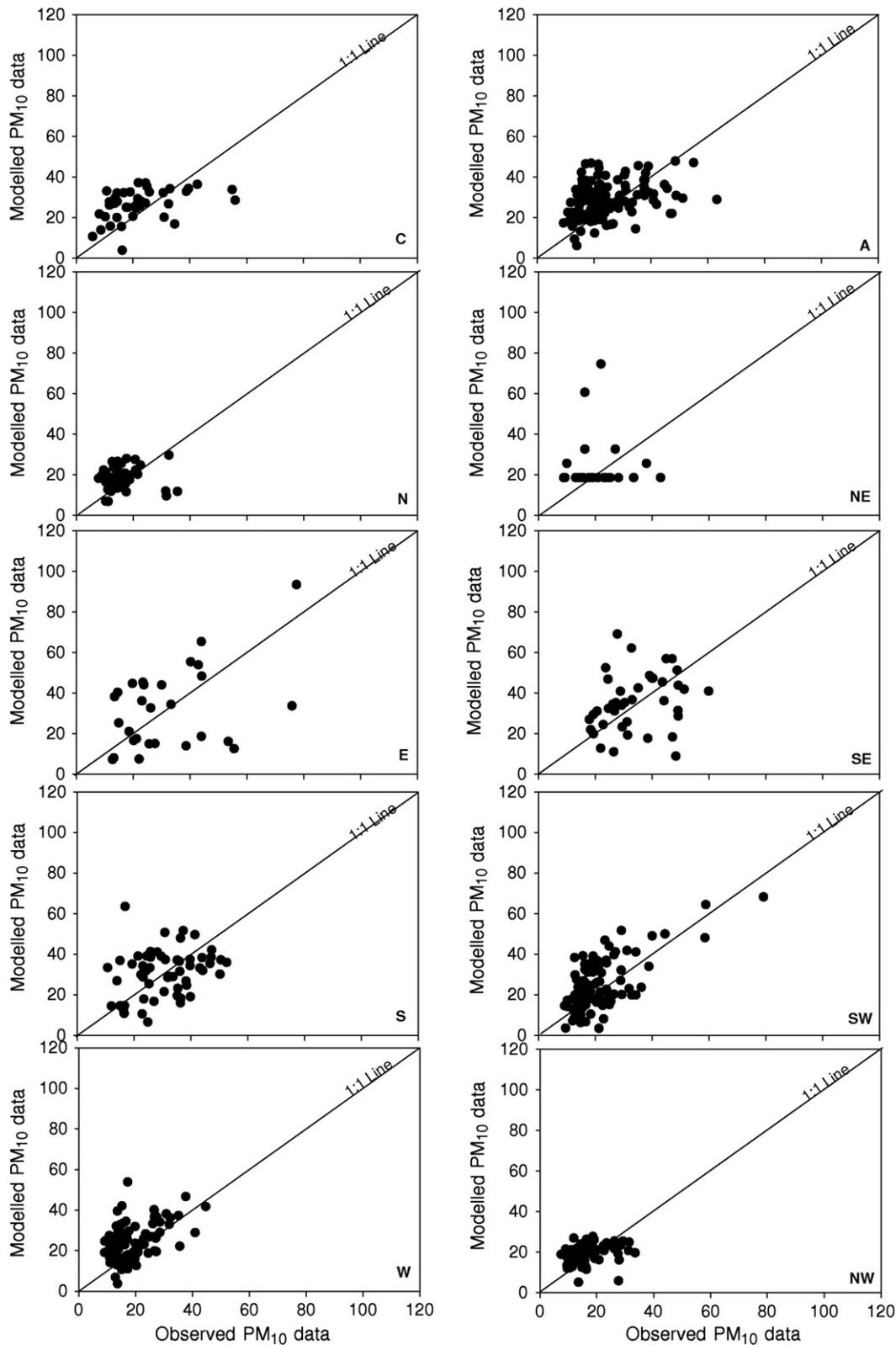
Fig. 6. Scatter plots showing observed versus modelled PM_{10} concentrations using observed meteorological data as input to the LWT_{MLR} model for each Lamb weather type for the whole year.

observed m8O$_3$ and PM$_{10}$ concentrations are analyzed for each weather type class separately (Figures 4–7).

High m8O$_3$ concentrations are restricted to the summer season (JJA), and more specific to the Lamb weather types A, E, SE and S. The modelled m8O$_3$ time series are in good agreement with the observed m8O$_3$ concentrations for the period 2005–2006 (Fig. 4). All distributions are thereby closely scattered around the perfect 1:1 line. When ECMWF data are used as predictors for the regression-based model, the highest concentrations are found for the same weather type classes (A, E, SE and S). However, the distributions are
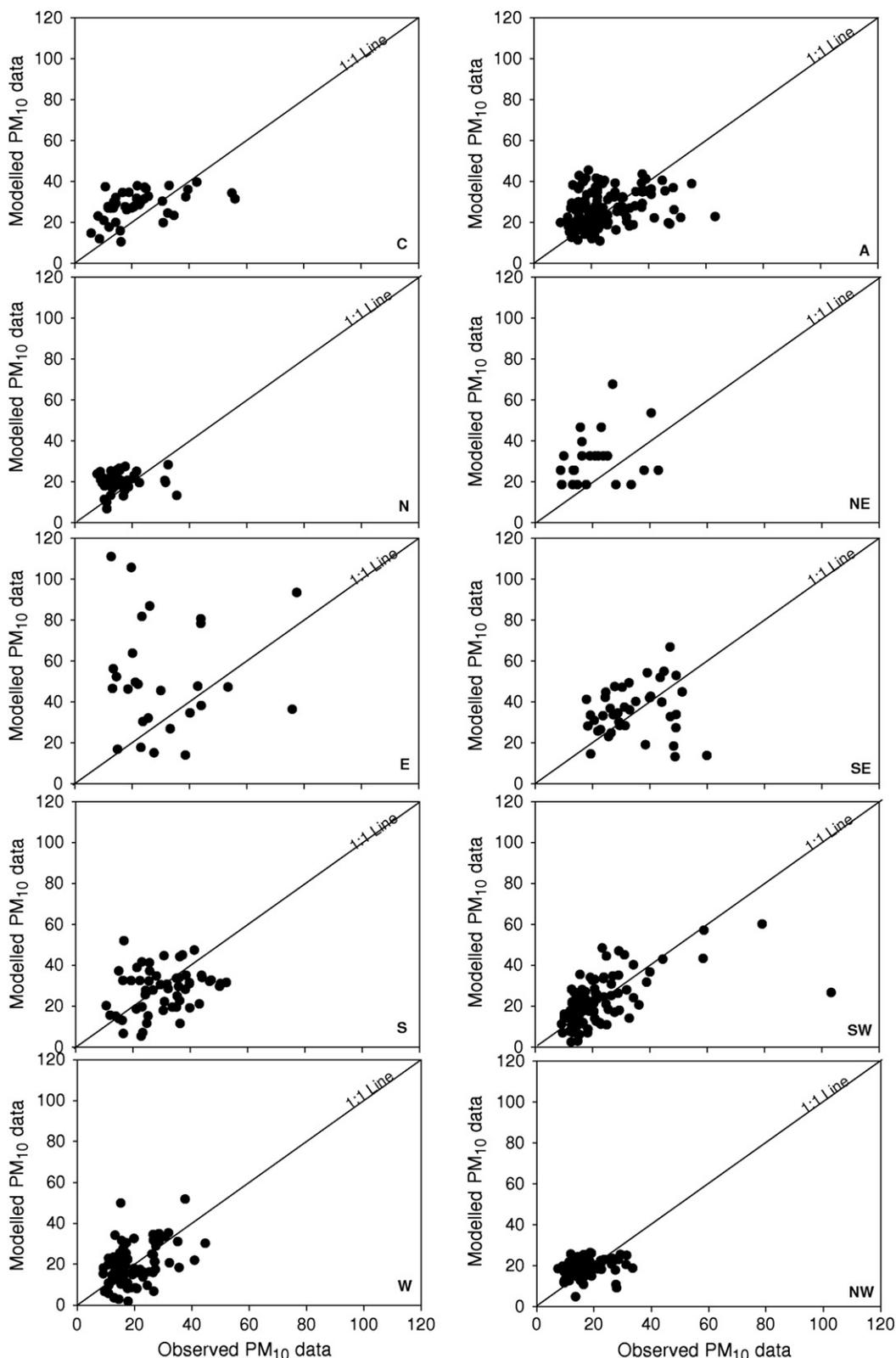
Fig. 7. Same as Fig. 6, but now using ECMWF meteorological data as input to the LWT$_{MLR}$ model.

characterized by a larger scatter and a lower explained variability, in the higher quantiles of the distribution (Fig. 5).

The modelled versus observed PM_{10} values show a large scatter around the optimal 1:1 line, especially for those types characterized by the highest concentrations (all types except N, NE, W and NW) both when observations and when ECMWF data are used as predictors (Figs. 6 and 7 respectively). Again, these results show that for PM_{10}, a regression-based approach using only meteorology data is insufficient to explain the observed PM_{10} variability. Therefore, the remaining of this sensitivity study of downscaled air-quality levels to its meteorological predictor values is restricted to $m8O_3$.

In order to understand the deficiencies of the regression-based model, the differences between ECMWF data and observations are studied only for the meteorological predictors that are included in the circulation pattern dependent regression equations (Fig. 8). Data are shown for the JJA months only, as this season is characterized by the highest $m8O_3$ concentrations.

Modelled $m8O_3$ values for the weather types A, SE and S are positively related to T_{max} (Demuzere et al., 2009), which is underestimated in its median value by up to 3 K in the ECMWF model. CC, RH and P0 are negatively related to ozone concentrations for the southern circulation pattern. These variables are overestimated by the ECMWF model (Fig. 8) and dampen high ozone concentrations. The underestimation of modelled $m8O_3$ concentrations under high-pressure conditions (type A) are favored by an overestimation of the relative humidity in ECMWF compared to the observations, which is again a predictor in the regression model for O_3 formation.

5. Discussion

Deficiencies in the regression-based models can be due to 1) an inadequate representation of the predictor variables and 2) the fact that most downscaling methods tend to resolve only part of the total variance (Barrero et al., 2006). In the previous section we found substantial differences between ECMWF data and observed meteorological variables. This difference can be due to 1) deficiencies in the ECMWF model or 2) the low spatial resolution used in our analysis. After all, in order to test our method for possible future downscaling purposes using AOGCM-output as an input for the regression-based models, we averaged 4 surrounding grid points from the spectral T63 horizontal grid, being representative for an area of about 350 × 350 km^2. By using such a coarse resolution, predictor variability is possibly dampened. In order to test this, a comparison is made with ECMWF operational forecast data from a single grid point (52.505°N/4.95° E) extracted from the original N400 Gaussian grid resolution (hereafter referred to as ECMWF1). The ECMWF1 data are representative for a much smaller area of about 25 × 25 km^2.

The quantile distribution of the ECMWF data is compared to the observed quantile distribution for the relevant meteorological variables (Fig. 9). The largest differences between ECMWF1 and ECMWF are found for T_{max}, RH and CC. T_{max} shows a large decrease in RMSE in ECMWF1 caused by a better representation of the higher percentile values. Although the RMSE for RH and CC derived from ECMWF1 is slightly higher than from ECMWF, the quantile distribution of ECMWF1 corresponds more closely to the observations. The impact of these differences in ECMWF predictor values on
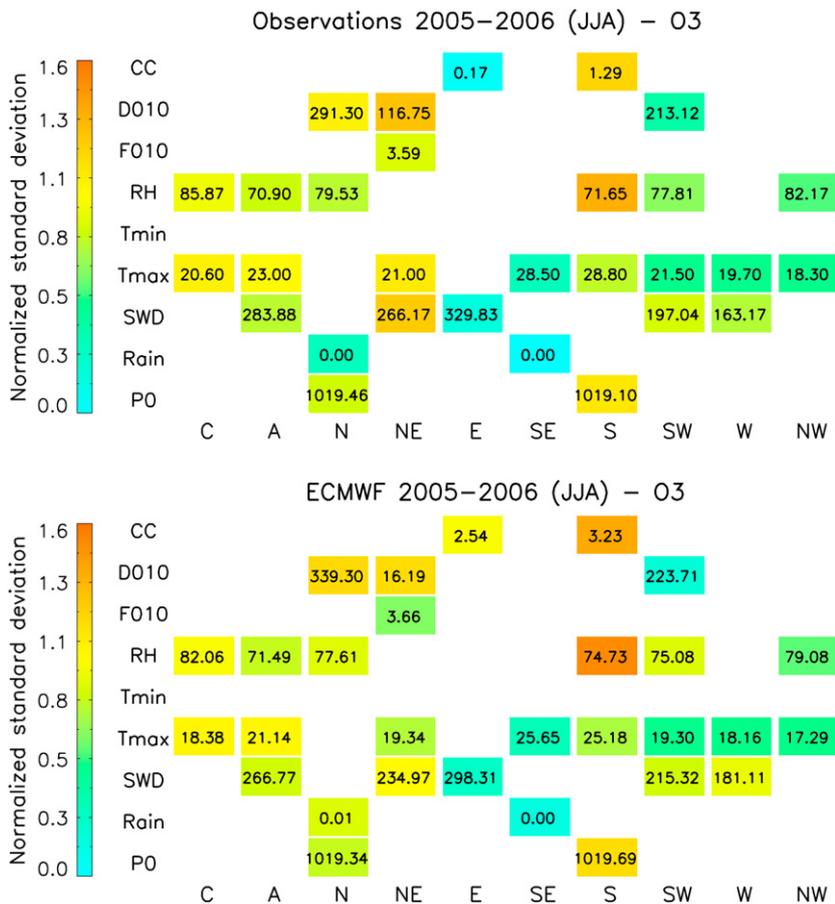
Fig. 8. Median (number) and normalized standard deviation (colors) for each of the meteorological predictor variables that are used in the circulation-specific regression equations, averaged over the summer (JJA) period 2005–2006. The upper panel presents observations, the lower panel present the results based on ECMWF data. (For interpretation of the colour plots and legend, the reader is referred to the web version of this article.)
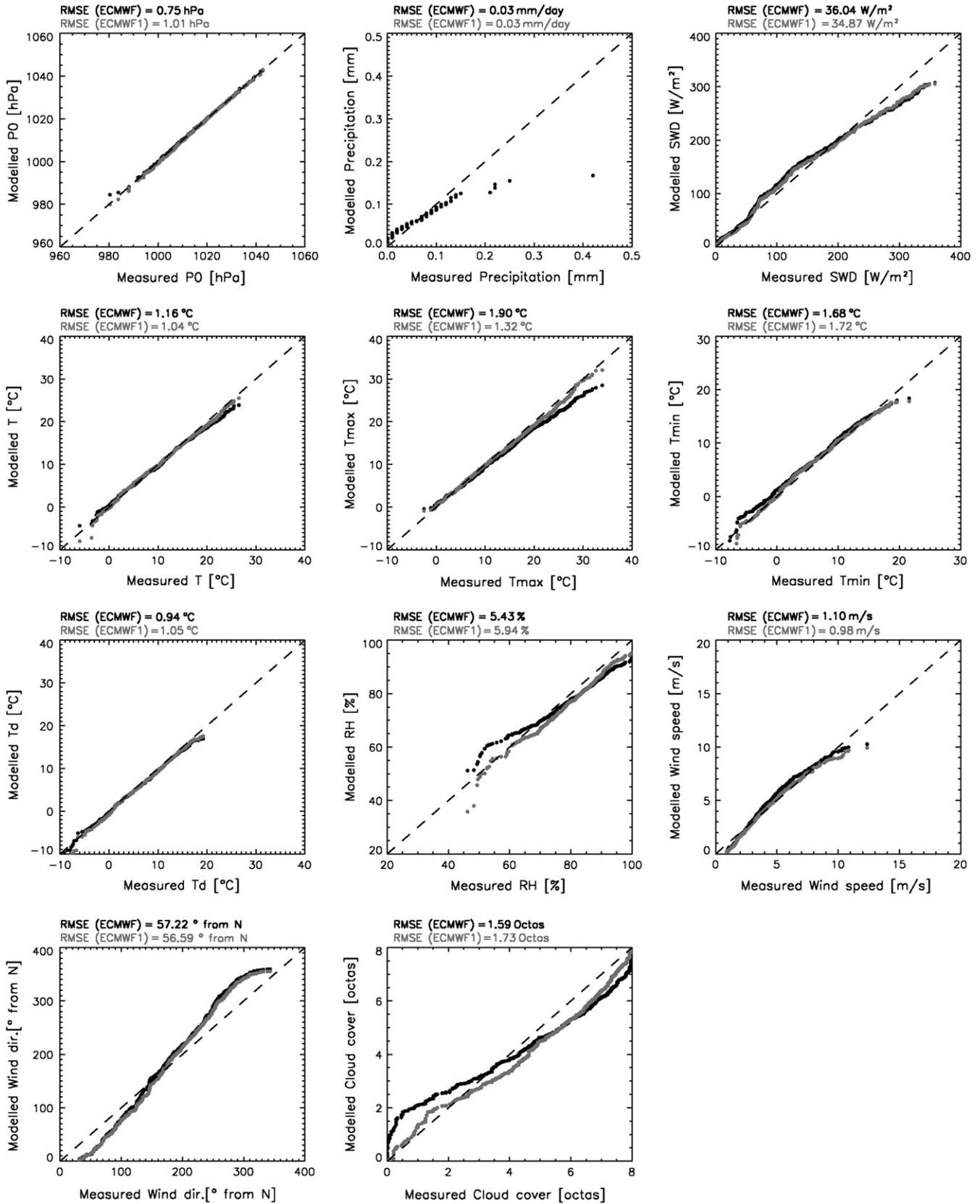
Fig. 9. Quantile–quantile plots of ECMWF versus observed meteorological predictor variables for the evaluation period 2005–2006. The 1:1 dashed line presents perfect ECMWF data and the RMSE present the error for both ECMWF spatial averaged on over an area of 350 km^2 (ECMWF) and ECMWF for one grid point (25 km^2) (ECMWF1) compared to reference observed meteorological values from Cabauw.

downscaled m8O$_3$ levels is plotted in terms of RMSE for LWT$_{SEAS+MLR}$ (Fig. 10). Furthermore, a qq-plot is used to present the distribution for both ECMWF and ECMWF1 downscaled m8O$_3$ concentrations using the LWT$_{SEAS+MLR}$ approach (Fig. 11). This comparison shows that in term of RMSE, there is a negligible difference for the downscaled m8O$_3$ time series using either spatial averages or single grid point predictor data. Nevertheless, the qq-plot (Fig. 11) shows that higher quantiles of observed m8O$_3$ concentrations (>120 μg m^{-3}) are better reproduced by the single grid point predictor data compared to the 4-grid point spatial average. This suggests that the differences between observed and ECMWF downscaled m8O$_3$ time series are both due to the usage of the 4-grid point average ECMWF data, dampening some of the predictors variability, as well as deficiencies in the ECMWF, resulting in an insufficient representation of the meteorological variables used as predictors in the regression-based model. Consequently, this result implies that downscaling techniques using low-resolution gridded data from for e.g. AOGCMs suffer from the low-resolution itself although improvement can already be obtained by addressing and carefully taking into account model errors, as was suggested by Cheng et al. (2007b).

The analysis in this manuscript is based on one single rural measurement site in the Netherlands. One could question the representativeness of this research to other sites in the world. Demuzere et al. (2009) showed that four different rural measurement stations in The Netherlands have similar characteristics, both in terms of the annual cycle of O$_3$ and PM$_{10}$ as in their relation with the suite of selected meteorological variables. This provides confidence in the spatial homogeneous character of rural sites for a mid-latitude area (in the Netherlands) and is in agreement with the results of Flemming et al. (2005). Nevertheless, this approach does not automatically imply a possible use in other non-rural environments, as the latter are more complex in terms of meteorological conditions (e.g. urban heat islands, street ventilation…) and local emission sources. As an example, ozone concentration in Belgian urban areas are usually lower than in their rural counterparts; such a 'VOC-limited' ozone formation regime is also observed in other North-Western and Central European urban areas (Hooyberghs et al., 2005). A first possible future approach to extend our analysis to non-rural areas could exist of the identification of different air-quality regimes for different pollutants, as suggested by Flemming et al. (2005) and
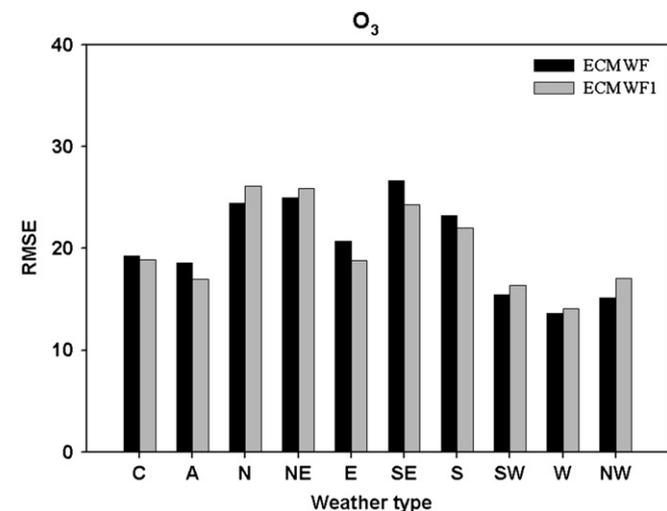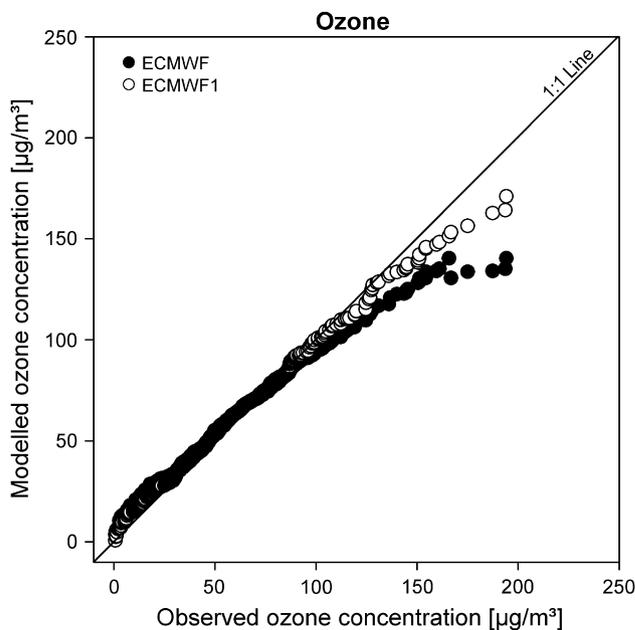
Fig. 11. Quantile–quantile plot of ECMWF and ECMWF1 downscaled O$_3$ versus observed O$_3$ concentrations for the evaluation period 2005–2006. The 1:1 line presents perfect model line.

secondly to develop a new set of circulation-dependant regression equations for these air-quality regimes.

6. Conclusion

The primary aim of this paper is to evaluate a variety of regression-based methodologies to hindcast levels of m8O$_3$ and PM$_{10}$ from meteorological predictors. In order to quantify the performance of the regression-based methods, several statistical indices are used besides the common first and second order moments: fit between modelled and observed series using the explained variance, shape of the distribution in terms of skewness and kurtosis and the performance against a persistence reference model. The analysis based on the calibration period 2001–2004 and an independent evaluation period 2005–2006 reveals that a stratification of the dataset using the automated Lamb weather type scheme, prior to the regression analysis improves the downscaling results for m8O$_3$ and PM$_{10}$ in terms of explained variance and skill score against the persistence model. Furthermore, simple model regressions have shown not to be able to capture the deviation from normality, e.g. with a non-zero skewness and kurtosis for the observed m8O$_3$ and PM$_{10}$ distributions. The introduction of a classification approach can reproduce these nonlinearity characteristics. Moreover, as m8O$_3$ is highly dependent on seasonal changes in its relations with meteorological predictors, the m8O$_3$ regression model is run in a seasonal mode, which is not helpful for PM$_{10}$.

Before the regression-based methods can be used for downscaling air-quality levels from coarse AOGCM-output, it needs to be tested whether the observed local relations between meteorological and air-quality variables hold on a larger scale. Therefore all regression-based models are evaluated using low-resolution ECMWF data interpolated on the spectral T63 horizontal grid for the period 2005–2006. Using these data as an input of the regression-based models leads to a slight decrease of the explained variance due to a lower variability of the meteorological predictors. Nevertheless, this approach can compete with other dynamical and statistical downscaling methods, which are often employed using observed time series, without being tested on low-resolution gridded data.

Fig. 10. Root mean square error (RMSE) values for O$_3$ for each weather type class derived from 4-grid point averaged ECMWF and single grid point ECMWF (ECMWF1) data using LWT$_{SEAS+MLR}$.

The somewhat deteriorated model performance is closely related to both ECMWF model deficiencies as well as to the coarse resolution of the data by averaging 4 surrounding grid points from the spectral T63 horizontal grid. The results from the regression-based model are similar in terms of RMSE, but better for the higher $m8O_3$ percentiles when using original ECMWF data on a Gaussian N400 resolution (corresponding to about 25 km^2) from the single grid point nearest to Cabauw as input to the regression-based model. Furthermore, comparing ECMWF meteorological data with observed data has shown some circulation-specific biases for these variables that play an important role in the ozone formation process. Therefore, we can conclude that apart from the dampening of variance due to the coarse resolution, ECMWF model deficiencies also control the lower performance compared to modelled $m8O_3$ and PM_{10} time series when observational meteorological data are used as predictors. It is clear that the errors in the predictor variables need to be reduced or quantified in order to produce realistic air-quality projections as a downscaling product of AOGCMs.

Our analysis shows that the use of a circulation approach prior to a linear regression model is beneficial for representing higher-order statistical moments of the air-quality level distributions and that this method is therefore suitable for downscaling O_3 concentrations. For PM_{10}, the situation is different as a regression-based approach using only meteorology data was found to be insufficient to explain the observed PM_{10} variability. Therefore, the second part of this analysis will only deal with projections of future $m8O_3$ concentrations derived from an AOGCM.

Acknowledgments

References

Ainslie, B., Steyn, D.G., 2007. Spatiotemporal trends in episodic ozone pollution in the lower fraser valley, British Columbia, in relation to mesoscale atmospheric circulation patterns and emissions. Journal of Applied Meteorology and Climatology 46, 1631–1644.

Barrero, M.A., Grimalt, J.O., Canton, L., 2006. Prediction of daily ozone concentration maxima in the urban atmosphere. Chemometrics and Intelligent Laboratory Systems 80, 67–76.

Cavazos, T., 1999. Large-scale circulation anomalies conducive to extreme precipitation events and derivation of daily rainfall in northeastern Mexico and southeastern Texas. Journal of Climate 12, 1506–1523.

Chaloulakou, A., Grivas, G., Spyrellis, N., 2003. Neural network and multiple regression models for PM10 prediction in Athens: a comparative assessment. Journal of the Air & Waste Management Association 53, 1183–1190.

Cheng, C.S.Q., Campbell, M., Li, Q., Li, G.L., Auld, H., Day, N., Pengelly, D., Gingrich, S., Yap, D., 2007a. A synoptic climatological approach to assess climatic impact on air quality in south-central Canada. Part I: historical analysis. Water Air and Soil Pollution 182, 131–148.

Cheng, C.S.Q., Campbell, M., Li, Q., Li, G.L., Auld, H., Day, N., Pengelly, D., Gingrich, S., Yap, D., 2007b. A synoptic climatological approach to assess climatic impact on air quality in south-central Canada. Part II: future estimates. Water Air and Soil Pollution 182, 117–130.

Cobourn, W.G., Dolcine, L., French, M., Hubbard, M.C., 2000. A comparison of nonlinear regression and neural network models for ground-level ozone forecasting. Journal of the Air & Waste Management Association 50, 1999–2009.

Cobourn, W.G., 2007. Accuracy and reliability of an automated air quality forecast system for ozone in seven Kentucky metropolitan areas. Atmospheric Environment 41, 5863–5875.

Comrie, A.C., Yarnal, B., 1992. Relationships between synoptic-scale atmospheric circulation and ozone concentrations in metropolitan Pittsburgh, Pennsylvania. Atmospheric Environment Part. B – Urban Atmosphere 26, 301–312.

Davies, T.D., Kelly, P.M., Low, P.S., Pierce, C.E., 1992. Surface ozone concentrations in Europe – links with the regional-scale atmospheric circulation. Journal of Geophysical Research – Atmospheres 97, 9819–9832.

Demuzere, M., Werner, M., van Lipzig, N.P.M., Roeckner, E., 2008. An analysis of present and future ECHAM5 pressure fields using a classification of circulation patterns. International Journal of Climatology 29.

Demuzere, M., Trigo, R.M., Vila-Guerau de Arellano, J., van Lipzig, N.P.M., 2009. The impact of weather and atmospheric circulation on O_3 and PM_{10} levels at a mid-latitude site. Atmospheric Chemistry and Physics 9, 2695–2714.

EU, 2008. Directive 2008/50/EC of the European Parliament and of the council of 21 May 2008 on ambient air quality and cleaner air for Europe. Official Journal of the European Union L 152/1.

Flemming, J., Stern, R., Yamartino, R.J., 2005. A new air quality regime classification scheme for O_3, NO_2, SO_2 and PM_{10} observations sites. Atmospheric Environment 39, 6121–6129.

Gachon, P., Dibike, Y., 2007. Temperature change signals in northern Canada: convergence of statistical downscaling results using two driving GCMs. International Journal of Climatology 27, 1623–1641.

Giorgi, F., Meleux, F., 2007. Modelling the regional effects of climate change on air quality. Comptes Rendus Geoscience 339, 721–733.

Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., Brasseur, O., 2005. A neural network forecast for daily average PM_{10} concentrations in Belgium. Atmospheric Environment 39, 3279–3289.

Huth, R., Kliegrova, S., Metelka, L., 2008a. Non-linearity in statistical downscaling: does it bring an improvement for daily temperature in Europe? International Journal of Climatology 28, 465–477.

Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, K., Tveito, O.E., 2008b. Classifications of atmospheric circulation patterns: recent advances and applications (in Trends and Directions in Climate Research). Ann. N.Y. Acad. Sci. 1146, 105–152.

IPCC, 2001. Climate change 2001: synthesis report. In: Watson, R.T., the Core Writing Team (Eds.), A Contribution of Working Groups I, II, and III to the Third Assessment Report of the Integovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom, and New York, NY, USA, p. 398.

Jones, P.D., Hulme, M., Briffa, K.R., 1993. A comparison of Lamb circulation types with an objective classification scheme. International Journal of Climatology 13, 655–663.

Marsland, S.J., Haak, H., Jungclaus, J.H., Latif, M., Roske, F., 2003. The Max-Planck-Institute global ocean/sea ice model with orthogonal curvilinear coordinates. Ocean Modelling 5, 91–127.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationship to the correlation coefficient. Monthly Weather Review 116, 2417–2424.

Roeckner, E., Bäuml, G., Boneventura, L., Brokopf, R., Esch, M., Girogetta, M., Hagemann, S., Kirchner, I., Kornblueh, L., Manzini, E., Schlese, U., Schulzweida, U., Tompkins, A., 2003. The Atmospheric General Circulation Model ECHAM5 Part I: Model Description. Report no. 354, Hamburg (Germany).

Tarasova, O.A., Brenninkmeijer, C.A.M., Jockel, P., Zvyagintsev, A.M., Kuznetsov, G.I., 2007. A climatology of surface ozone in the extra tropics: cluster analysis of observations and model results. Atmospheric Chemistry and Physics 7, 6099–6117.

Thode, H., 2002. Testing for Normality. Marcel Dekker, New York.

Wilby, R.L., Wigley, T.M.L., 1997. Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography 21, 530–548.

Wilby, R.L., Charles, S.P., Zorita, E., Timbal, B., Whetton, P., Mearns, L.O., 2004. Guidelines for use of climate scenarios developed from the statistical downscaling methods, Available from the DDC of IPCC TGCIA, 27 pp. Available from: IPCC-DDC. http://www.ipcc-data.org.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego.

Willmott, C.J., 1981. On the validation of models. Physical Geography 2 (2), 184–194.

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 30, 79–82.