Final report:

Air mass origin and its influence over the parameters measured at Cabauw Tall Tower. Influence of the NAO index on the flow pattern

María Cabello Ganuza

Elche (Spain), 8 November 2011

1. Done work during the stay at the ECN:

The main aim of this work was to identify the main flows arriving at Cabauw (51.971º N, 4.927º E), and relate them with the recorded immission levels of greenhouse gases. Likewise, the influence of the NAO index variation on the atmospheric transport was also evaluated in several Atlantic and Mediterranean locations. To reach the objectives during my 3-months stay, the following tasks have been done:

- **Retrieval of the ERA-Interim meteorological data** with a resolution of 1º for the period 1993-2009. These data have been used to run both the HYSPLIT and FLEXTRA transport models
- **Computation of back-trajectories** using two different models (HYSPLIT and FLEXTRA) and three different meteorological datasets (2.5º NCEP/NCAR Reanalysis, and 1º and 1.5º ERA-Interim). Finally, four back-trajectory data sets were obtained (Table 1). 96-hour backward air trajectories arriving at 0, 6, 12 and 18 UTC at 200, 500, 1500 and 3000masl at Cabauw for the period 1993-2009 were computed.
  In addition, 96-hour back-trajectories arriving at 12 UTC at Cabauw, Mace Head, Elche, Lecce and Málaga at 3000, 1500 and 500masl for the period 1990-2009 were computed using HYSPLIT transport model and 1.5º ERA-Interim meteorological data for the subsequent analysis of their relationship with the "station" NAO index.

Table 1: Back-trajectory data sets used in this study.

Variable	h-RP	h-ERA	h-ERAInt	f-ERAInt
Meteo-data	NCEP/NCAR Reanalysis Project	ERA-Interim	ERA-Interim	ERA-Interim
ATM*	HYSPLIT	HYSPLIT	HYSPLIT	FLEXTRA
Horizontal resolution	2.5º lat-long	1.5º lat-long	1º lat-long	1º lat-long
Vertical resolution	17 levels from 1000 to 10 hPa	60 levels with the model top at 50 hPa	60 levels with the model top at 0.1 hPa	60 levels with the model top at 0.1 hPa
Time interval	6-hourly	3-hourly	3-hourly	3-hourly

*ATM: Atmospheric Transport Model

- **Computation of differences between the trajectories** obtained by both different transport models and meteorological data. The great-circle distance was used.
- **Back-trajectory cluster analysis** using the k-means cluster method with the great-circle distance. Methodology is explained in subsection 2.3.
- **Characterization of the main flows** as identified by the cluster analysis and study of their **relationship with immission levels of greenhouse gases** measured in the period 1993-1997 and 2000-2009 at the study site. To detect statistically significant differences in CH_4 and CO_2 measurements at 200masl at Cabauw Tall Tower according to the identified clusters, the Kruskal-Wallis and the pairwise Mann-Whitney tests (with the Dunn-Sidak correction) have been used.
- Study of the **relationship between the variation of the NAO index and the variation of the frequency of the main flows** at five different locations: two Atlantic sites (Cabauw and Mace Head), and three Mediterranean sites (Málaga, Elche and Lecce).

2. Results

The influence of using different atmospheric transport models and/or different meteorology and resolution dataset was evaluated for the trajectories arriving at Cabauw (Fig. 1). Moreover, a clustering analysis was carried out and related with the greenhouse gases immisions. Finally, a study regarding the relationship with the NAO index and the air flow pattern was done.

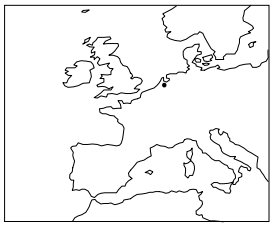

Fig. 1: Location of the Cabauw Tall Tower.

2.1 Horizontal great-circle distance between trajectories

Different works can be found in the literature which deal with trajectory errors (Rolph and Draxler, 1990; Stohl *et al*., 1995; Harris *et al*., 2005); although there exists some confusion regarding the statistical parameters used. The measure that has been adopted by many authors in recent years and also in this study is the horizontal transport deviation (HTD):

$$HTD(t) = \frac{1}{N}\sum_{n=1}^{N} dist_n(t)$$

where *N* is the number of trajectory pairs to compare, and $dist_n(t)$ is the great-circle distance between the two points corresponding to *t* hours of the *n*th pair of trajectories to compare. The accumulated horizontal transport deviation (AcHTD) is also considered computed by summing the HTD values every hour up to *t*.

The largest mean differences are found when comparing trajectories computed with different model, resolution and meteorological data (i.e. h-RP *vs* f-ERAInt). In turn, the lowest distances are found for differences due to the use of different model (h-ERAInt *vs* f-ERAInt) (Fig. 2).

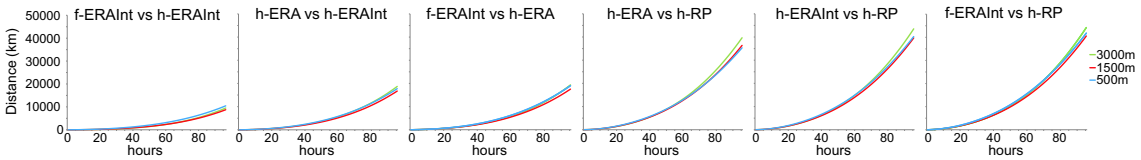

Fig. 2: Accumulated mean horizontal transport deviation, AcHTD, when comparing trajectories computed using different data input arriving at 500 (blue line), 1500 (red) and 3000masl (green).

In general, for the first days differences are larger for the trajectories arriving at 500m and then at 3000m. Figure 2 shows also that the largest distances are found at 3000m in all the cases except when comparing the use of different models (f-ERAInt vs ERAInt) where it is at 500m. This could be associated to higher winds at higher altitudes. In turn, the models use, among other things, different orography which has more influence at low heights. On the other hand, the trajectories are more similar at 1500m.

The AcHTD(t) between trajectories presents a good quadratic fit in all the cases (Fig. 3), as Cabello *et al*. (2008) found in their study to assess the influence of meteorological datasets on the trajectories.

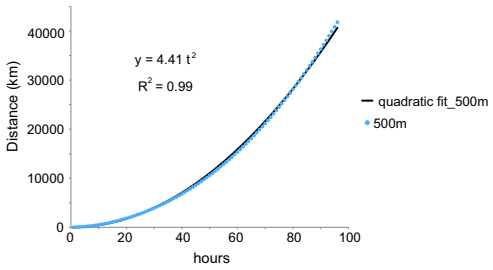

Fig. 3: AcHTD between f-ERAInt and h-RP trajectories arriving at the study site at 500m (blue dots) and quadratic fit (black line).

The distribution of the $dist_n$ values computed for 96 hours backward, shows a log-normal distribution irrespective of the trajectories which are compared. This would imply that these differences are the result of the product of many small and random effects. The width of the distribution depends on the similarities or dissimilarities between the trajectories: The log-normal distribution is narrower and sharper (Fig. 4, left) when comparing trajectories computed with different atmospheric transport model (h-ERAInt vs f-ERAInt) which means that trajectories are very close. In turn, when comparing trajectories computed with different ATM, resolution and meteorological transport (f-ERAInt vs h-RP) the log-normal distribution is wider and has a lower peak (Fig. 4, right).

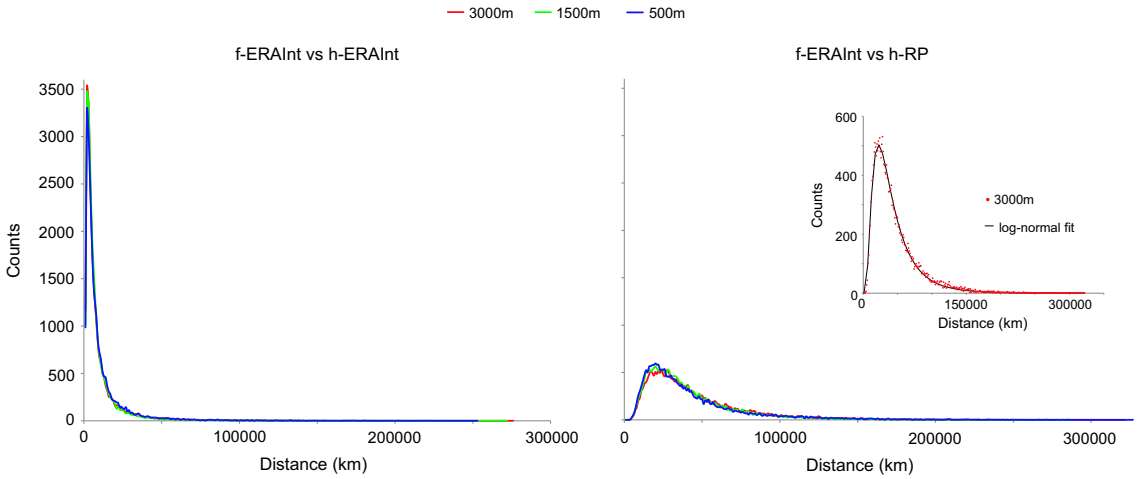

Fig. 4: Density distribution of $dist_n$ at 96 hours for comparison of back-trajectories arriving at 3000 (red), 1500 (green) and 500m (blue line) computed using different ATM (left), and different ATM, resolution and meteorological data (right). Density distribution of the $dist_n$ at 96 hours and log-normal fit for trajectories arriving at 3000m and computed using different ATM, resolution and meteorological data (inset).

The HTD presents a strong seasonality, with greater distances in wintertime probably due to the faster velocities of the air masses (Fig. 5). The highest altitude the largest differences, although the lowest distances between trajectories computed at different heights are found in summertime. Like in Figure 2, the smallest distances are found when comparing trajectories computed using different atmospheric transport model (f-ERAInt vs h-ERAInt), followed by trajectories calculated with different resolution (h-ERAInt vs h-ERA), and the largest differences are found when comparing different atmospheric transport model, different resolution and different meteorological data (f-ERAInt vs h-RP) (Fig. 5).

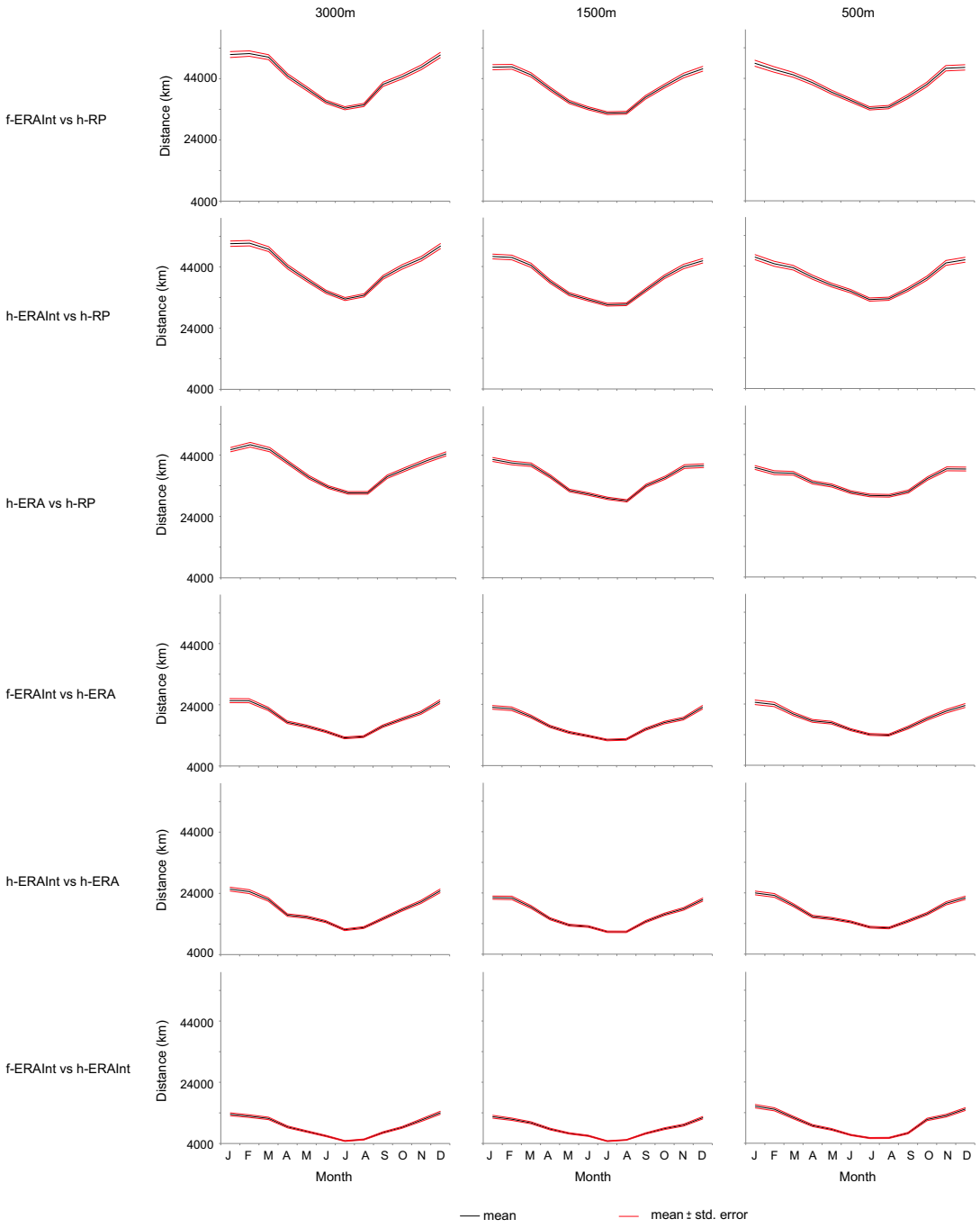

Fig. 5 HTD seasonality for trajectories arriving at Cabauw at 3000m (left), 1500 and 500m (right).

2.2 Horizontal great-circle distance and meteorological scenario

The pathway of an air mass is defined by the evolution of the synoptic situation and hence by a meteorological scenario. Small differences in height or location might lead to dramatic differences in trajectory pathways. When studying the days when the distances between trajectories are the biggest ones, a relationship with complex meteorological situations, mostly the presence of an occluded front (Fig. 6b), is found. In fact, the occlusion process is not well understood, and different explanation of the atmospheric dynamics when forming an occluded front is found in the literature (Schultz and Mass, 1993; Martin, 1998) pointing out the complexity of this kind of process.

Figure 6 shows two different meteorological scenarios. The first one is dominated by a high pressure which drives the pathway of the trajectories both calculated with Flextra model and ERA-Interim meteorological data, and with HYSPLIT model and Reanalysis meteorological data (Fig. 6a,c). In the second one, a cold front is overtaking a warm front and hence, an occluded front is forming in the area where trajectories computed by f-ERAInt and h-ERA became separated. This occurs within the first day, crossing Belgium.

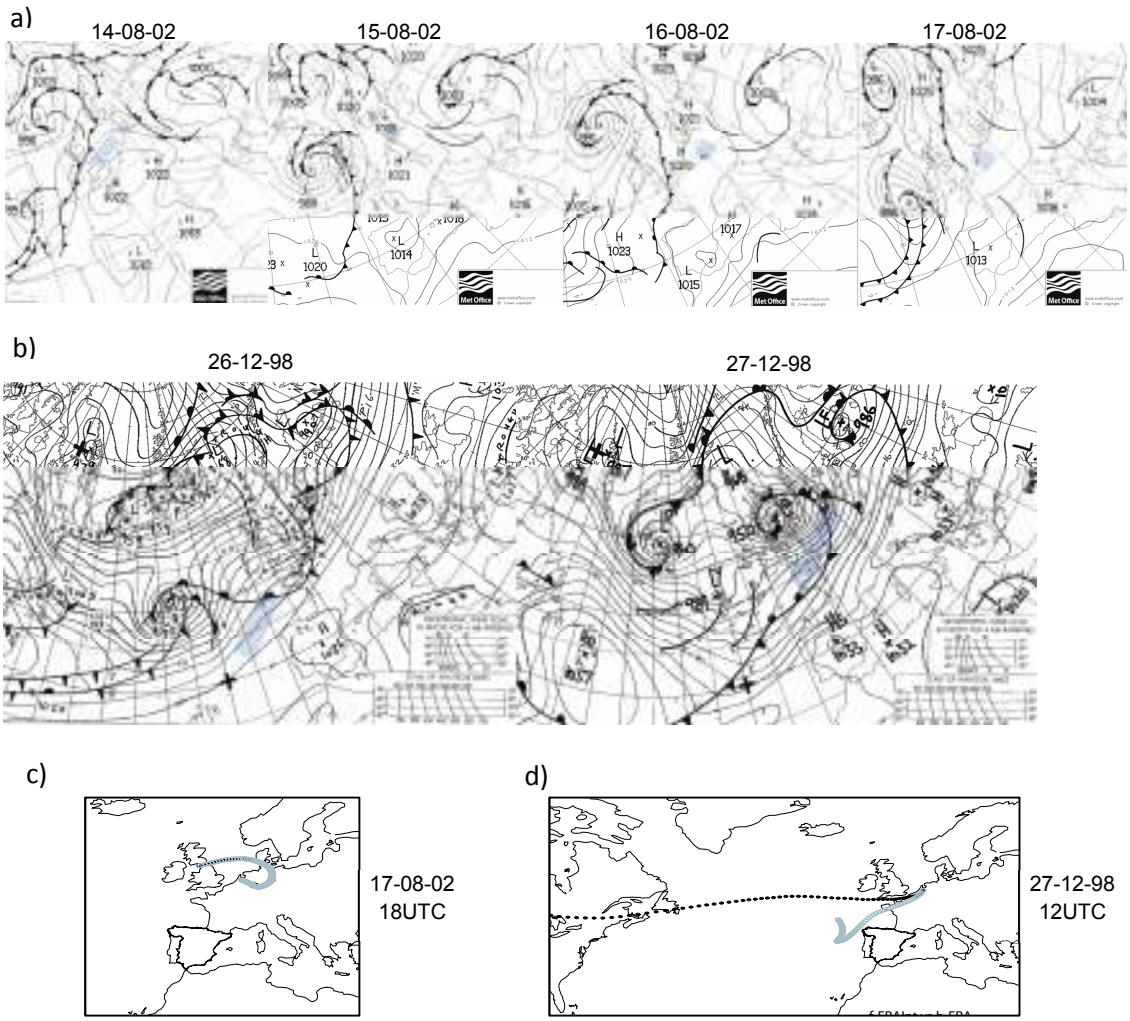

Fig. 6: 00UTC synoptic charts where fronts are depicted, for days with (a) the closest and (b) the most distant trajectories arriving at 500m. The former corresponds to (c) August 17, 2002 for the trajectories computed at 18UTC by f-ERAInt (grey circles) and h-RP (black dots). The latter corresponds to (d) December 27, 1998 for the trajectories computed at 12UTC by f-ERAInt (black dots) and h-ERA (grey circles). Grey boxes show the trajectory pathway within that day.

2.3 Clustering

The clustering of back-trajectories has been made using the k-means method with the great-circle distance. To reduce the subjectivity in the selection of the appropriate number of clusters, 800 replicate solutions are computed for each number of clusters (*k* ranges from 30 to 3) and the one with the smallest total within-group sum of squared distances (RMSD) is selected. Afterwards, from a plot of RMSD by number of clusters, we retain the smallest number of clusters for which the smallest total RMSD change is found (Cabello et al., 2008).

The cluster analysis of the 6-hourly back-trajectories arriving at Cabauw at 3000, 1500 and 500m groups the main air flows into 6 air flow types (N, W, NWfast, WSW, SW and E) with independence of the meteorological data or model used (Fig. 7). The largest groups in number of trajectories are W and SW accounting together for more than 40% of the total; however, the smallest groups are NWfast and WSW which account together around 25% of the total.

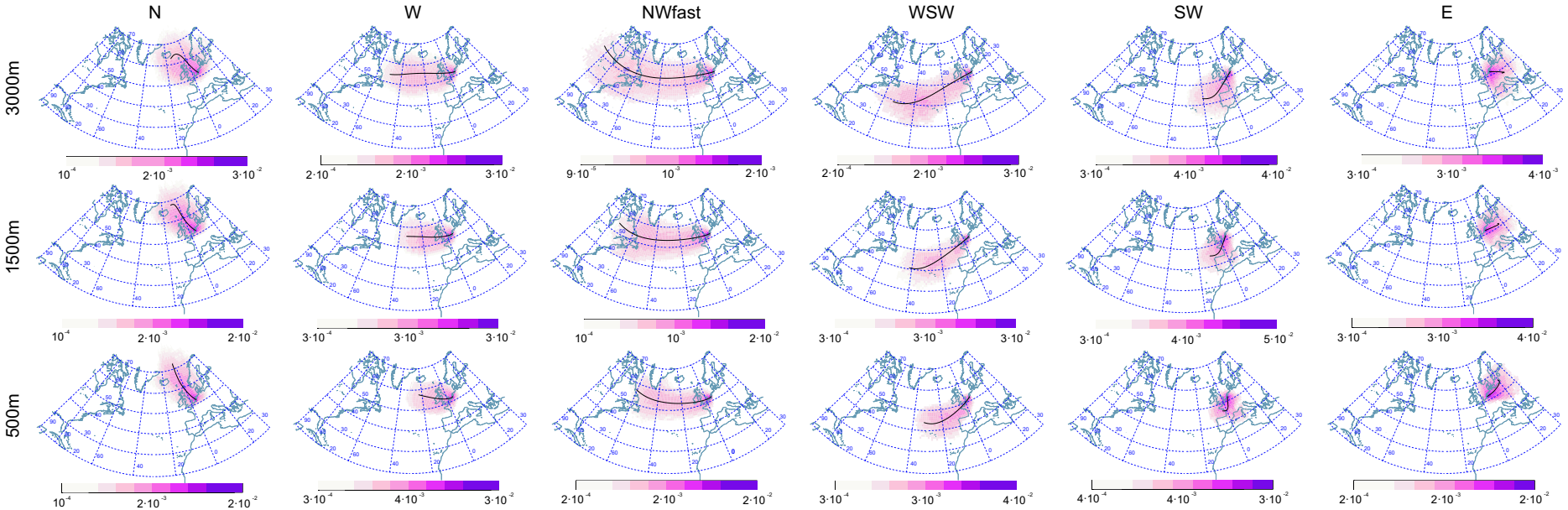

Fig. 7: Residence time relative to the total number of endpoints within every cluster, for the trajectories calculated using the Flextra model with ERAInterim meteorological data (f-ERAInt), for each identified air flow type. Black line shows the centroid of the cluster.

Most of the trajectories correspond to westerly flows (W, NWfast, WSW, SW), more than 65% of the total. The trajectories arriving in summertime belong mainly to W (at all the heights), and SW at 3000 and 1500m. The winter trajectories correspond mainly to NWfast flows. In spring, the main flows are coming from the East at all the heights, and from the North at the lowest altitudes. Finally, the trajectories arriving at 500m from the SW are mainly found in autumn (Fig. 8).

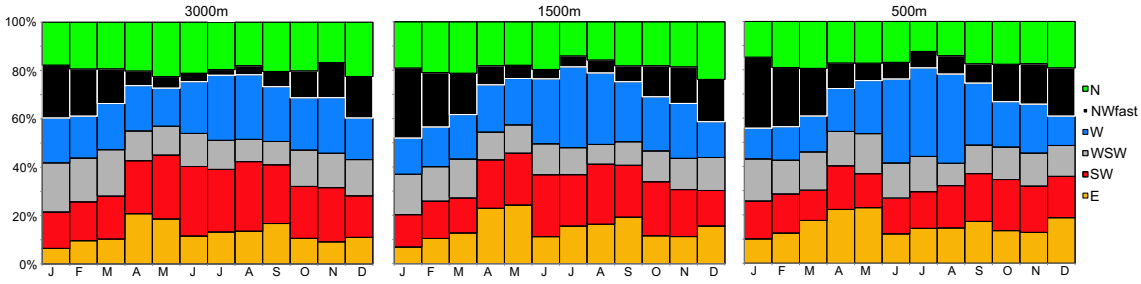

Fig. 8: Seasonality for the trajectories arriving at the study site at 3000, 1500 and 500m computed using Flextra model and ERA-Interim meteorological data (f-ERAInt).

The differences between centroids corresponding to similar air flow type, computed by different clusterings (different trajectory sets), are larger with decreasing arrival height. There is an exception for trajectories arriving from SW, where the highest distance is found at 1500m. The closest centroids correspond to N trajectories (1423.40 and 6443.69km at 3000 and 1500m respectively), except at 500m which belong to SW (3877.90km). The most different centroids correspond to trajectories arriving from NWfast (13916.40, 14315.69 and 18321.13km at 3000, 1500 and 500m respectively). The HTD for 96 hours between centroids shows thousands kilometers of distances, but the pathway of every cluster is similar with independence of the ATM and/or meteorological data chosen (Fig. 9).

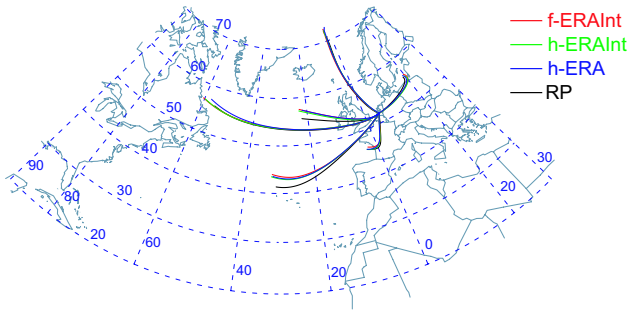

Fig. 9: Centroids at 500m for the clustering with different trajectory sets.

The number of trajectories which are assigned to the same flow-type when using different meteorological data and atmospheric transport model, is a maximum for E and N flows. In turn, the least coincident group is W. The percentage of coincidence ranges between more than 90% when comparing f-ERAInt vs h-ERAInt to around 60% when comparing f-ERAInt vs h-RP (Table 2).

Table 2: Percentage of trajectories classified in the same type of air flow when considering different clustering procedures, both for the most similar classification (f-ERAInt vs h-ERAInt) and the most dissimilar one (f-ERAInt vs h-ERAInt).

f-ERAInt vs h-RP

		E	N	NWfast	SW	W	WSW
3000m	E	79	8	0	13	0	0
	N	5	79	1	4	12	0
	NWfast	0	1	78	0	10	11
	SW	5	3	0	78	7	6
	W	0	8	7	7	70	8
	WSW	0	0	10	6	8	75
1500m	E	80	6	0	13	1	0
	N	7	81	3	2	7	0
	NWfast	0	3	78	0	7	12
	SW	7	2	0	69	15	5
	W	0	11	10	5	69	4
	WSW	0	0	8	7	13	72
500m	E	85	7	0	5	2	0
	N	6	81	5	1	8	1
	NWfast	0	2	71	2	7	18
	SW	14	0	0	71	10	4
	W	3	10	8	7	67	5
	WSW	1	0	7	16	16	61

f-ERAInt vs h-ERAInt

		E	N	NWfast	SW	W	WSW
3000m	E	96	2	0	2	0	0
	N	1	96	0	0	3	0
	NWfast	0	0	93	0	4	3
	SW	2	1	0	95	1	1
	W	0	2	2	2	93	2
	WSW	0	0	2	1	2	95
1500m	E	94	2	0	4	0	0
	N	1	96	1	0	2	0
	NWfast	0	0	95	0	2	2
	SW	1	0	0	95	2	1
	W	0	2	2	2	93	1
	WSW	0	0	2	1	2	95
500m	E	95	2	0	0	1	0
	N	1	95	1	0	2	0
	NWfast	0	1	93	0	1	5
	SW	4	0	0	93	2	1
	W	1	2	2	1	92	2
	WSW	0	0	2	4	4	90

2.4 Air masses and its relationship with immission levels of greenhouse gases

The statistical analysis to detect significant differences in both CH_4 and CO_2 measurements at 200m at Cabauw Tall Tower according to the identified clusters shows that the immission of greenhouse gases are well differed by the main flows. The lowest CH_4 concentrations mainly correspond to NWfast flows (3000 and 1500m), and to N flows at 500m. The least CO_2 polluted flows are those coming from W and N (only at lower height). On the other hand, the most polluted air masses are the E (3000m) and SW (500m) for both pollutants (Table 3). This would be the expected result as long as the main polluted sources are located in mainland (Europe) and the sea acts as a sink.

Table 3: Mean and standard deviation of CH_4 and CO_2 measurements for the identified clusters for the trajectories computed by f-ERAInt. Equal number in the superscript indicates no significant differences between values as stated from the Mann-Whitney test. The highest and lowest values are highlighted in red and black, respectively.

3000m	CH_4 (mean±std dev)	CO_2 (mean± std dev)
NWfast	**1901±58**[1]	380±14[2]
W	**1910±72**[1,2]	**379±14**[1]
SW	1962±103[4]	**383±17**[3,4]
N	1914±74[2]	380±16[2]
WSW	1916±67[3]	382±14[3]
E	**1985±121**[5]	**384±17**[4]

1500m	CH_4 (mean± std dev)	CO_2 (mean± std dev)
NWfast	**1899±53**[1]	381±14[3]
W	1913±73[2]	**379±15**[1]
SW	1975±113[4]	**384±17**[4]
N	**1902±65**[1]	380±15[2]
WSW	1915±67[3]	382±14[3]
E	**1981±109**[5]	**384±17**[4]

500m	CH_4 (mean± std dev)	CO_2 (mean± std dev)
NWfast	1895±49[2]	380±14[2]
W	1916±71[3]	**378±15**[1]
SW	**1995±118**[5]	**387±18**[4]
N	**1895±59**[1]	**378±14**[1]
WSW	1910±65[3]	381±14[2]
E	1975±103[4]	384±16[3]

It is noteworthy that the origin of the trajectories at the different heights is not decoupled unlike the trajectories at the western Mediterranean where they are strongly decoupled (Cabello *et al.*, 2008). In fact, the most frequent combinations of the three heights are SW_{3000m}-SW_{1500m}-SW_{500m}; $E3_{000m}$-E_{1500m}-E_{500m}; N_{3000m}-N_{1500m}-N_{500m} and W_{3000m}-W_{1500m}-W_{500m}, each combination accounting for more than 5% of the total. It is interesting to emphasize that the two first combinations correspond also to days with CH_4 concentrations larger than the third quartile; in turn, the most frequent combination with CO_2 concentration larger than the third quartile is beneath 2% of the total.

In addition, a preliminary study of the goodness of the modeled pathways is done using the COMET model (Vermeulen *et al.*, 2006). The CO_2 concentrations estimated by this Lagrangian model have been compared to the ones measured at 20, 60, 120 and 200masl at the Cabauw Tower. Three input datasets are used: f-ERAInt, h-ERAInt and h-RP, for trajectories arriving at 200masl at Cabauw from May to July of 2008. As Table 4 shows, better correlations are found with decreasing heights.

Table 4: R^2 values from the correlation between the estimated and the measured CO_2 concentrations at the Cabauw Tall Tower.

Input data	20m	60m	120m	200m
f-ERAInt	0.45	0.30	0.13	0.08
h-ERAInt	0.47	0.35	0.16	0.09
RP	0.44	0.35	0.14	0.08

2.5 NAO index (NAOi)

The study of the interannual variations in tropospheric transport pathways in southern Europe and its relationship to the North Atlantic Oscillation (NAO) is made for five locations: two Atlantic stations (Cabauw and Mace Head) and three Mediterranean stations (Málaga, Elche and Lecce) (Fig. 10).

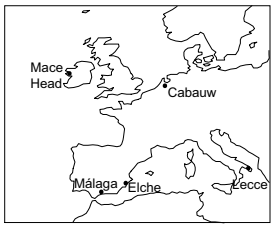

Fig. 10: Location of the five study sites for the relationship of the variations in NAOi and the frequency of occurrence of the main flows.

To assess the year-to-year meteorological variability in the Northern Hemisphere a 20-year period (1990-2009) was considered. A cluster analysis of back-trajectories arriving at 12UTC at 3000, 1500 and 500m at each study site was made by using the same procedure described above. The 96-hours back-trajectories were computed with the HYSPLIT model v4.8. ERA-Interim data with resolution of 1.5º lat-long was used as input; 50 hPa was the top of the model.

The cluster analysis groups the main air flows arriving at 3000, 1500 and 500m respectively into 5, 6 and 6 at Mace Head; 6, 6 and 6 at Cabauw; 7, 5 and 6 at Málaga; 5, 5 and 4 at Elche and 7, 7 and 6 at Lecce (see Fig. 11). Most of the trajectories (more than 50%) correspond to westerly flows identified as northwesterlies (NWfast/NWmod), westerlies (W/Wslow) or westerly-southwesterly (WSW) in all the study sites, as should be expected. In general, air masses show a clear seasonal pattern (Table 5). Northwesterly flows are more frequent during the wintertime, while flows arriving from the East are common in spring or summertime.

Table 5: Seasonality of the identified clusters for every location. A cluster belongs to a season when has assigned more trajectories than the third quartile (trajectories$_k$-Q_{3k}>0).

	Height	Spring	Summer	Autumn	Winter	No seasonality
Mace Head	**3000m**		NWslow		NWfast	NWmod SW E
	1500m	N	W		NWfast SW	WSW E
	500m	N	SW W		NWfast	WSW E
Cabauw	**3000m**	E	SW W		NWfast WSW	N
	1500m	N E	SW W		NWfast WSW	
	500m	E	W	SW	NWfast WSW	N
Málaga	**3000m**	NWmod N	SW W S		NWfast Wfast	
	1500m	N	SW E		NWfast W	
	500m		NWmod E		NWfast Wfast	N W
Elche	**3000m**	NWmod	SW E		NWfast W	
	1500m	NWmod	E Wslow	W	NWfast	
	500m	N	E		NW	W
Lecce	**3000m**		NE W	SW	NWfast	NWmod
	1500m		NE W	SW WSW	NWfast NWmod	Wfast
	500m	NE WSW	W	SW	NWfast	NWmod

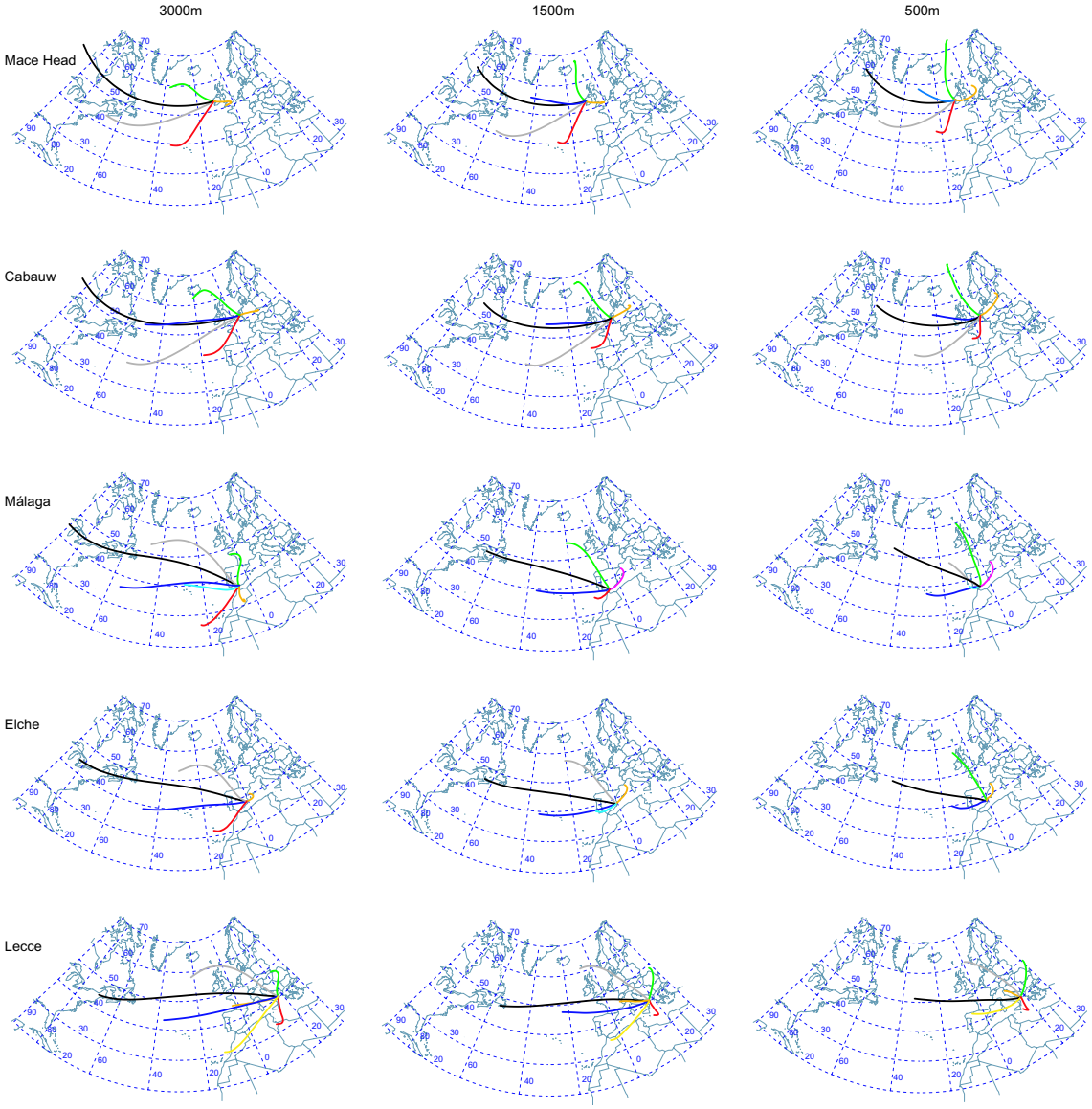

Fig. 11: Centroids for the identified air flow types arriving at 3000, 1500 and 500masl at each location.

The relationship of the monthly NAOi and the frequency of occurrence of each identified air flow type, for winter months (from December to March), shows in some cases a linear regression with slope significantly different from zero ($p<0.01$) (Table 6). In the Atlantic locations there is a strong increase in polar maritime air masses of relatively high wind speed (NWfast, WSW) as well as a decrease of E flows in the positive NAO phase. Such an increase in NWfast flows is also registered in the central Mediterranean. In the western Mediterranean, however, there is no increase of NWfast flows, and easterlies are more frequent in the NAO+ phase. Moreover, W flows are less frequent in the positive phase in this area in the lower troposphere (1500, 500 masl). In short, the association to the NAO is much higher at the Atlantic locations than in the Mediterranean.

Moreover, when NAO phases are depicted, a clearly pattern is shown (both in the Atlantic façade and the Mediterranean basin)(Fig. 12). For the NAO+ phase, the westerly and fast pathways are predominant; in turn, the NAO- phase is related to more frequent southwesterly and slow flows.

Table 6: Significance of the slope from the relationship between the monthly NAOi and the occurrence of the identified air flows. Crosses mean that the slope is not significantly different from zero ($p<0.01$). Positive/negative symbol shows the sign of the slope. No symbol means that the flow type is not found at that height.

		N	NW	NWfast	NWmod	W	Wfast	SW	S	E
Málaga	**3000m**	X	X	X		X	X	-	+	
	1500m	X		X		-		X		+
	500m	X		X	X	X	-			X

		N	NW	NWfast	NWmod	W	Wslow	SW	E
Elche	**3000m**			X	X	X		X	X
	1500m			X	X	-	X		X
	500m	X	X			-			X

		NE	NWfast	NWmod	W	Wfast	WSW	SW
Lecce	**3000m**	X	+	X	X	X	-	X
	1500m	X	+	X	X	X	-	X
	500m	X	+	X	X		-	X

		N	NWfast	W	WSW	SW	E
Cabauw	**3000m**	-	+	X	+	-	-
	1500m	-	+	X	+	-	-
	500m	X	+	X	+	-	-

		N	NWfast	NWmod	NWslow	W	WSW	SW	E
Mace Head	**3000m**		+	+	-			X	-
	1500m	-	+			X	+	X	-
	500m	-	+			X	+	X	-

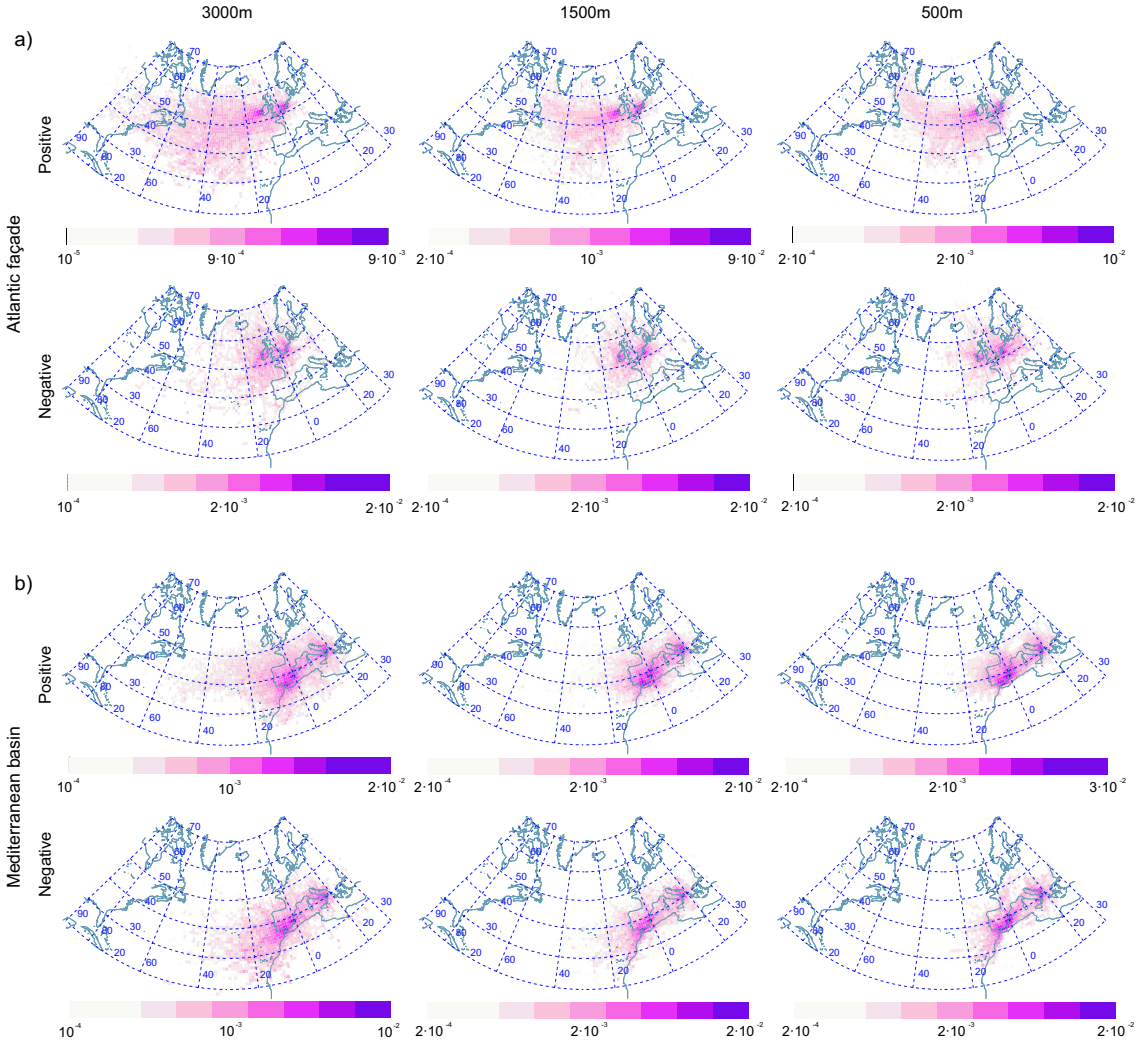

Fig. 12: Residence time relative to the total number of endpoints for the study sites located at (a) the Atlantic façade and (b) the Mediterranean basin for the NAO+ and NAO- phases. Only wintertime data is used.

3. Conclusions

The study of the trajectory sensitivity to the atmospheric transport model and meteorological dataset has been carried out. Distances between trajectories computed using different model are the smallest ones, while the largest ones are found when using different model and meteorological data. These differences are larger for wintertime.

Dramatic differences might be found in some event, normally caused by a complex meteorology like occluded fronts. However, when performing trajectory cluster analysis, differences caused by the use of different model or meteorological dataset, are almost negligible.

The NAO+ phase leads to westerly and fast pathways, while southwerterly and slow pathways are associated with the NAO- phase. This relation with the NAOi is higher at the Atlantic locations than in the Mediterranean basin.

References:

Cabello, M., Orza, J.A.G., Galiano, V. and Ruíz, G.: Influence of the meteorological datasets on the backtrajectory cluster analysis. A 7-year study in SE Spain, Adv. Sci. Res.,2, 65-70, 2008.

Harris, J.M., Draxler, R.R. and Oltmans, S.J.: Trajectory model sensitivity to differences in input data and vertical transport method, J. Geophys. Res. 110, D14109, doi:10.1029/2004JD005750, 2005.

Martin, J.E.: The structure and evolution of a continental winter cyclone. Part I: Frontal structure and the occlusion process, Monthly weather review, 126, 303-328, 1998.

Rolph, G.D. and Draxler, R.R.: Sensitivity of three-dimensional trajectories to the spatial and temporal densities of the wind field, J. Appl. Meteorol. 29, 1043-1054, 1990.

Schultz, D.M. and Mass, C.F.: The occlusion process in a midlatitude cyclone over land, Monthly Weather Rev., 121, Issue 4, 918-940, 1933.

Stohl, A., Wotawa, G., Seibert, P., Kromp-Kolb, H.: Interpolation errors in wind fields as a function of spatial and temporal resolution and their impact on different types of kinematic trajectories. J. Appl. Meteorol. 34, 2149-2165, 1995.

Vermeulen, A.T., Pieterse, G., Hensen, A., van den Bulk, W.C.M. and Erisman, J.W.: Comet: a lagrangian transport model for greenhouse gas emission estimation forward model technique and performance for methane, Atmos. Chem. Phys. Discuss., 6, 8727-8779, 2006.