THEME ENV.2012.6.1-1

EUPORIAS

(Grant Agreement 308291)

European Provision Of Regional Impact Assessment on a

Seasonal-to-decadal timescale

Deliverable 33.3

Report describing formulation of strategies for communicating confidence levels for S2D forecasts

Deliverable Title	D33.3 Report describing formulation of strategies for communicating confidence levels for S2D forecasts
Brief Description	D33.3 Report describing formulation of strategies for communicating confidence levels for S2D forecasts: Report will provide a sample of visualisations of existing and new approaches for visualising confidence levels. This sample will match the potential use by stakeholders of S2D forecasts
WP number	33
Lead Beneficiary	University of Leeds
Contributors	**WP33 partners** Andrea Taylor, University of Leeds Suraje Dessai, University of Leeds Carlo Buontempo, Met Office Mike Butts, DHI Laurent Dubus, EDF Ghislain Dubois, TEC Eroteida Sánchez, AEMET Christian Viel, Meteo France **WP32 contribution** Jesus Fernandez, University of Cantabria Maria Dolores Frias, University of Cantabria
Creation Date Version Number Version Date	02/03/2015 1.2 30/04/2015
Deliverable Due Date Actual Delivery Date	30/04/2015 30/04/2015
Nature of the	x R - Report

Deliverable		
		P - Prototype
		D - Demonstrator
		O - Other
Dissemination Level/ Audience	x	PU - Public
		PP - Restricted to other programme participants, including the Commission services
		RE - Restricted to a group specified by the consortium, including the Commission services
		CO - Confidential, only for members of the consortium, including the Commission services

Version	Date	Modified by	Comments
V1.1	02/03/2015	Andrea Taylor	
V1.2	30/04/2015	Andrea Taylor	Also edited by Jesus Fernandez and Maria Dolores Frias

EUPORIAS

Table of Contents

List of Figures

1. Executive Summary

In this report we detail the strategies for communicating uncertainty and confidence developed in Task 33.3. Strategy development was informed by both our review of the literature on communicating uncertainty (T33.2) and our survey of user needs (T33.1), which demonstrated both that user's vary in their ability to utilise statistical and technical information, and that forecast skill is not always communicated to users in a way that is easily interpreted. We have therefore sought to identify and develop a portfolio of formats for communicating likelihood (uncertainty) and skill (confidence) to users varying in statistical and technical expertise. For expert users these formats include tercile plots, bubble plots (maps) incorporating measures of likelihood and skill, and measures of spread. For those without high statistical expertise we suggest formats that use evaluative categories and text to describe skill. Having developed these formats we set out a plan for testing a) how well these communications are understood; b) which formats are preferred by users; and c) how they would be used in decision making.

2. Project Objectives

With this deliverable, the project has contributed to the achievement of the following objectives (DOW, Section B1.1):

No.	Objective	Yes	No
1	Develop and deliver reliable and trusted impact prediction systems for a number of carefully selected case studies. These will provide working examples of end to end climate-to-impacts-decision making services operation on S2D timescales.		
2	Assess and document key knowledge gaps and vulnerabilities of important sectors (e.g., water, energy, health, transport, agriculture, tourism), along with the needs of specific users within these sectors, through close collaboration with project stakeholders.		
3	Develop a set of standard tools tailored to the needs of stakeholders for calibrating, downscaling, and modelling sector-specific impacts on S2D timescales.		
4	Develop techniques to map the meteorological variables from the prediction systems provided by the WMO GPCs (two of which (Met Office and MeteoFrance) are partners in the project) into variables which are directly relevant to the needs of specific stakeholders.		

5	Develop a knowledge-sharing protocol necessary to promote the use of these technologies. This will include making uncertain information fit into the decision support systems used by stakeholders to take decisions on the S2D horizon. This objective will place Europe at the forefront of the implementation of the GFCS, through the GFCS's ambitions to develop climate services research, a climate services information system and a user interface platform.	X	
6	Assess and document the current marketability of climate services in Europe and demonstrate how climate services on S2D time horizons can be made useful to end users.		

3. Detailed Report

3.1 Objectives

The overarching objective of this Work Package is to develop good practice in communicating levels of confidence and uncertainty in seasonal-to-decadal (S2D) climate predictions. Having identified a number of communication challenges faced by the providers and users of climate information in previous work package Tasks (Taylor & Dessai, 2014; Taylor et al., 2014); the aim of Task 33.3 has been to formulate strategies for communicating levels of confidence in S2D that would address these. In this report we briefly outline the key findings of Tasks 33.1 and 33.2 (section 3.2), before detailing our procedure for identifying and developing communication strategies to address the challenges identified in these task (3.3), and the formats emerging from this process (3.4). Finally we set out plans for testing the strategies developed (3.5).

3.2 Previous findings

In Task 33.1 we conducted a user needs survey with a sample of EUPORIAS stakeholders and representatives from other organisations expressing an interest in S2D prediction (Taylor & Dessai, 2014). This was followed in Task 33.2 by a review of existing approaches to communicating uncertainty, which incorporated both an examination of those formats currently being used to communicate uncertainty in the context of S2D, and a discussion of findings on the communication of uncertainty emerging from the literature on risk and decision making (Taylor et al., 2014).

3.2.1 User preferences

With respect to user preferences, the findings of our user-needs survey indicated that participants tended to favour maps and visualisations depicting forecast spread (e.g. error bars). Although in the latter case this preference was stronger amongst those who were more comfortable with using statistical information. A strong association between the existing familiarity of communication strategies and preference also emerged; a finding consistent with earlier work examining the communication of climate change projections

(Daron, 2014). Hence, users may be reluctant to use novel formats unless their usefulness and user-friendliness can be easily demonstrated. However, it should be kept mind that user's preferred formats may not always be those that are most well understood (Lorenz et al., 2015). For this reason one key goal of this task was to identify and develop both novel and familiar formats, with the goal of examining whether those formats that are most preferred are also most well understood in a subsequent Decision Lab (Task 33.4).

3.2.2 Differences between users

As noted above, our user needs survey showed that preference for particular formats is sometimes linked to comfort with statistical information. This is consistent with findings from the broader risk communication literature, which show that those who are less fluent at using statistical information may struggle to effectively interpret and utilise quantitative data (Peters et al., 2009), and suggests that non-experts may benefit from the presence of qualitative 'evaluative categories' to aid decisions (Peters et al., 2009, Gregory et al., 2012). We also found that respondent's organisations varied in their tolerance for uncertainty and false alarms: highlighting the importance of making both probability and forecast skill salient. As Stephens et al. (2012) point out however, in communicating uncertainty in climate information to users a trade-off can exist between *richness* (amount of detail), *robustness* (appropriate reflection of skill and forecast limitations) and *salience* (understandability). That is to say that richly detailed communications (e.g. those with fine grained temporal or spatial resolution) may not be supported by the forecast. While complex information about either forecast detail or forecast skill may not be well understood by all users; raising the risk that this information will be disregarded or misinterpreted (Kain and Covi, 2013). In selecting communications strategies for further development we have therefore endeavoured to choose both formats suitable for expert users, and formats suitable for users with less experience of using statistical information.

3.2.3 The challenge of integrating information about probability and skill

Another finding emerging from the user needs survey was that a large proportion of those who indicated that their organisation currently received S2D, indicated that they did not receive information about how well forecasts have performed in the past. As forecast performance is reflected by measures of skill and reliability, this suggests that that this information is either not being provided or it is not being clearly and understandably conveyed to many users.

Summary: Previous Findings

- End-users vary in expertise, meaning that the needs and preferences of those who are technical and statistical experts differ from those who are less familiar with statistics.

- When it comes representing uncertainty people tend to prefer familiar formats, but these may not always be those that are best understood. Testing is therefore needed to establish whether preference matches objective understanding.

- At present, information about forecast performance (i.e. skill and reliability) is not always provided to users, or at least not always provided in a way that is clear and understandable. It is important that this issue be addressed.

3.3 Procedure for identifying and developing communication strategies

The selection and development of the strategies for communicating uncertainty identified in this report was an iterative process, involving ongoing collaboration between WP33 partners and the University of Cantabria (WP32), who produced the visualisations and the accompanying R code. The selection process was also advised by visualisation specialist Aiden Slingsby (City University). As WP33 aims to complement the bespoke formats for communication being created for EUPORIAS prototypes, by developing formats that can be used in a broad range of contexts, it was established that the portfolio of communication strategies selected should fulfil the following criteria:

- Formats should be capable of appropriately representing uncertainty for a range of different climate variables and indices.
- It should be feasible for forecast providers to produce these formats for a range of different climate variables and indices.
- The set of strategies selected for development should contain both formats suitable for expert users with advanced knowledge of statistics, and formats suitable for non-experts who have limited experience of using statistical information.

3.4 Strategies for communicating levels of confidence and uncertainty in S2D predictions

The examples presented in 3.4.1 to 3.4.5 are provided by Jesus Fernandez and Maria Dolores Frias (University of Cantabria), and represent sample surface temperature data retrieved from ECOMS-UDG (https://meteo.unican.es/trac/wiki/udg/ecoms). Predictions are retrieved from System 4 (15 ensemble members) and observations from WFDEI (Weedon et al., 2014). The time periods considered for these plots is 1982 to 2010. Plots are for northern hemisphere winter (December to February), with one month lead time. The visualisations presented in this report depict forecasts for the Iberian Peninsula (area of lower skill) and Ethiopia (area of slightly greater skill). However, in the associated R code (see supplementary material at the end of this report to view the R code) the region of interest can be changed by altering the latitude and longitude specified. Also the season, lead time and other parameters can be adjusted. All plots except the tercile plot (3.4.1) show the forecast for a particular year (2010, in the examples).

It should be noted that, in the case of the Ethiopian forecasts represented in this report, climate change trends are leading to artificial skill; with cold events being clustered at the start of the time period considered and warm events towards the end (see Figures 1b and 1d for a visual illustration of this). This issue should be resolved in the future by the creation of detrended plots.

The examples presented in 3.4.6 are pilot formats developed as possible methods of providing information about likelihood and forecast skill to less statistically experienced users. As little research has previously examined how both likelihood and skill can most effectively be presented to this group, these will be pilot tested for understandability with a public sample and amended before being shown in the Decision Lab. The decision to use a temperature forecast in the development of these communication strategies was taken because user needs surveys have indicated that this variable is of interest to most respondent organisations (Taylor & Dessai, 2014). However, the formats developed may be easily adapted to depict uncertainty about other climate variables and indices (e.g. precipitation, drought indices, etc). In selecting the strategies for further development we have aimed to identify both formats capable of providing spatial information and formats providing temporal information.

In keeping with existing recommendations for the use of colour in climate visualisations (see for instance (Kaye et al., 2012), we have opted to utilise a red=warmest, grey=middle tercile, and blue=coolest scheme for representing terciles, as red and blue tend to be intuitive linked with hot and cold respectively. Owing to the prevalence of red-green colour-blindness amongst the general population, grey rather than green has been selected to represent the middle tercile. It should, of course, be kept in mind that this colour scheme may not be appropriate for other variables or indices where colour associations may be different (e.g. it would be counterintuitive to use blue to depict 'low precipitation'). However, this may be easily altered by changing the specifications within the R code.

In developing the formats below, one issue for consideration has been how best to represent forecast skill. That is to say whether to use a measure focusing on the occurrence (or not) of a given event (e.g. ROCSS of the exceedance of a given tercile; see Jolliffe & Stephenson, 2003), or a measure that reflects the skill of the forecasting system on a range of possible

outcomes (e.g. RPSS; see Jolliffe & Stephenson, 2003). As skill may vary between different quantiles, the former provides a detailed and accurate picture as to how much 'confidence' should be placed in the occurrence of a given quantile range. However, when presenting forecast information using a measure of spread such as a confidence interval or probability density function (pdf), the ROCSS cannot be so easily integrated with this information, since there is no binary forecast (occurrence/non occurrence). Likewise, the increase in complexity that comes with presenting a skill score for each quantile may render some formats especially difficult for less experience users to interpret (e.g. the bubble plot in Figure 3e). Hence, a trade-off between 'robustness' and 'salience' (understandability) may exist. Alternative versions of the visualisations outlined in 3.4.5 have been produced with ROCSS and RPSS respectively; and it is our intention to further examine how users respond to these in the Decision Lab.

RPSS has been selected as a summary of the skill of a system, integrating the skill in each tercile. ROCSS is a numerical summary of the ROC curve and we used it for binary probabilistic forecast (tercile occurrence). This is the only probabilistic numerical summary recommended by the WMO's Lead Centre for Long-Range Forecast Verification System (http://www.bom.gov.au/wmo/lrfvs).

3.4.1Tercile plot

The tercile plots in Figures1a-d depict the predicted probability of upper (warmer than average), middle (average) and lower (cooler than average) terciles. Skill represented using ROCSS for each tercile. The shading of each square represents probability (darker shade=greater probability), while ROCSS is presented numerically for each tercile. White dots indicate the tercile actually observed in each historical time period. Tercile plots are intended to show the performance of the forecast system along a time period. They can be also be modified to include an additional column with the forecast for the next season (without the dot representing the observed outcome, of course). This plot can be used to allow users to "get a sense" of what a particular ROCCS value means in terms of past performance of the probabilistic forecasts.

In our earlier user need's survey visualisations that depicted discrete categories in graph form were less popular amongst respondents than maps or measures of spread. However, this plot differs in that shading rather than area is used to represent probability. The plot also visually displays the extent to which past predictions have corresponded with observations: information that a sizable minority of respondents indicated that they did not currently receive but would like to.

Two versions of this plot are illustrated above: one using greyscale to represent probability of each tercile (a-b), the other using a separate colour for each tercile (c-d). Both versions have been identified as having particular benefits. The greyscale version (a-b) may make it easier to compare the shading across the three terciles, while the colour version (c-d) has the benefit of making it easier to distinguish which terciles represent 'warmer than average' and 'cooler than average'. The decision has been taken to present (c-d) in the Decision Lab. However, the code for both versions is available, and may be easily adjusted to meet colour preferences.

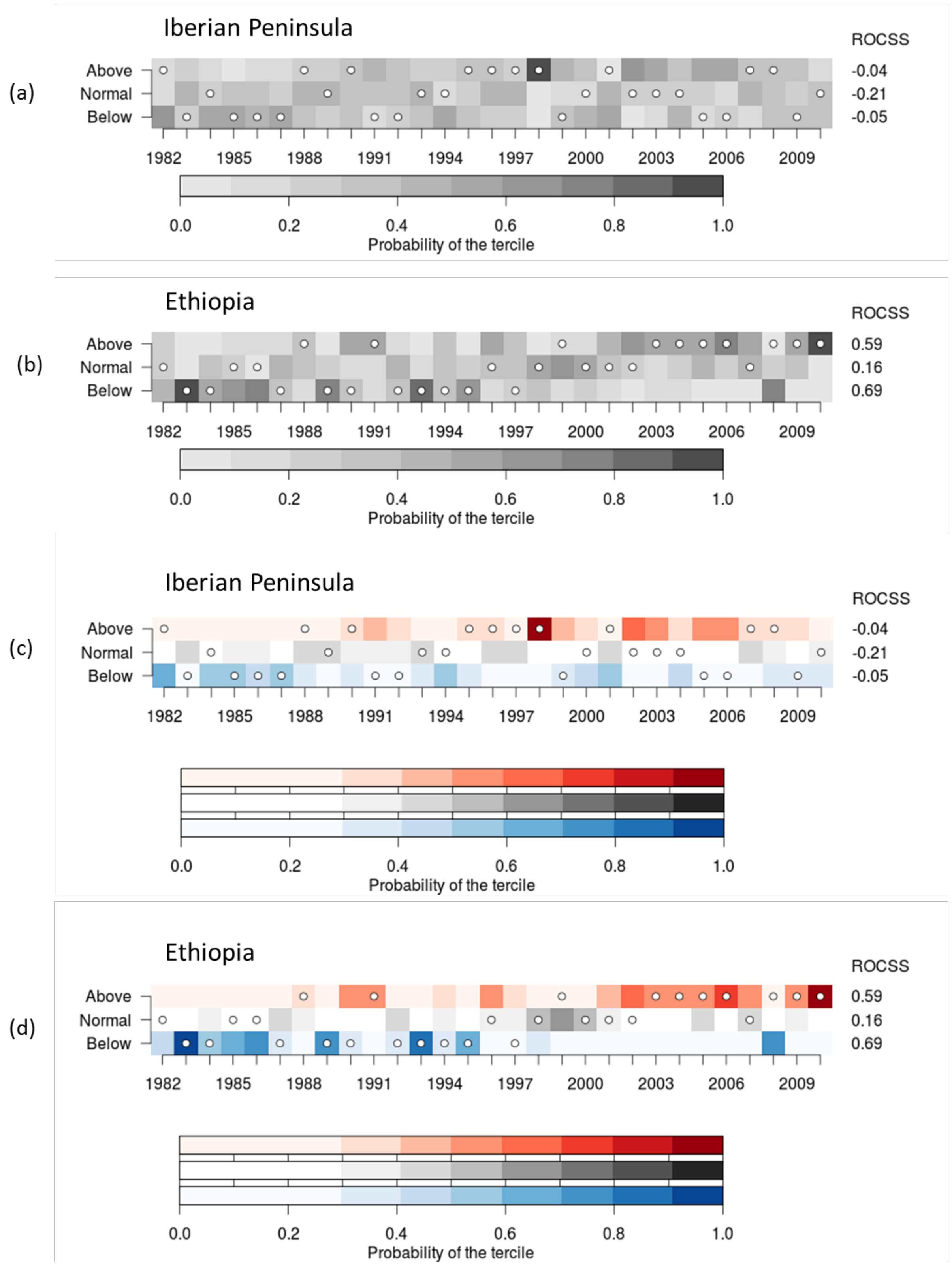


Figure 1a-d Tercile plots with lightness of colour representing predicted probability of above average, average, and below average temperatures for the Iberian Peninsula **(a, c)** and Ethiopia **(b, d)** for the time period 1982 – 2009. Plots **(a)** and **(b)** use greyscale only. Plots **(c)** and **(d)** use a different colour for each tercile. Skill score (ROCSS) is represented numerically for each tercile. White dots denote the observed tercile for each time period.

3.4.2 Tercile bar graph

This simple bar graph (Figures 2a-b) is based on the one currently used by MeteoSwiss in their online seasonal temperature and precipitation forecasts (http://www.meteoswiss.admin.ch/home/climate/future/seasonal-outlook.html). It depicts the predicted likelihood of upper, middle and lower tercile relative to climatology (represented by a grey line). Skill score (ROCSS) is presented for each tercile in numeric form. Skill score is colour coded to indicate whether there is No Skill (Red), Some Skill (Grey), or Good Skill (Blue); so as provide a salient warning to users where scores are negative.

A version of this visualisation where the shading of the bars was used to represent forecast skill has also been created. However, as this led to confusion over whether shading reflected probability or skill, the decision was taken to directly colour code numeric information about the ROCSS.

(a)
(b)

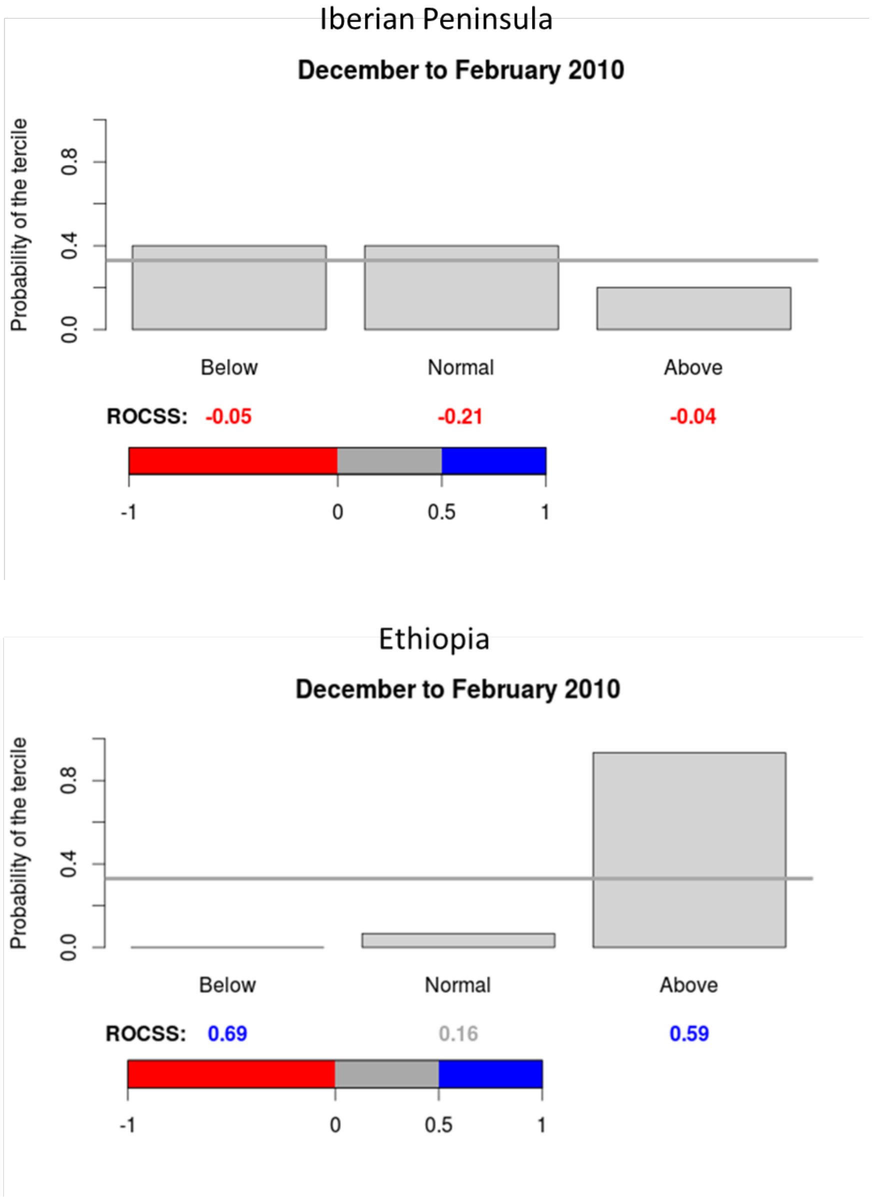


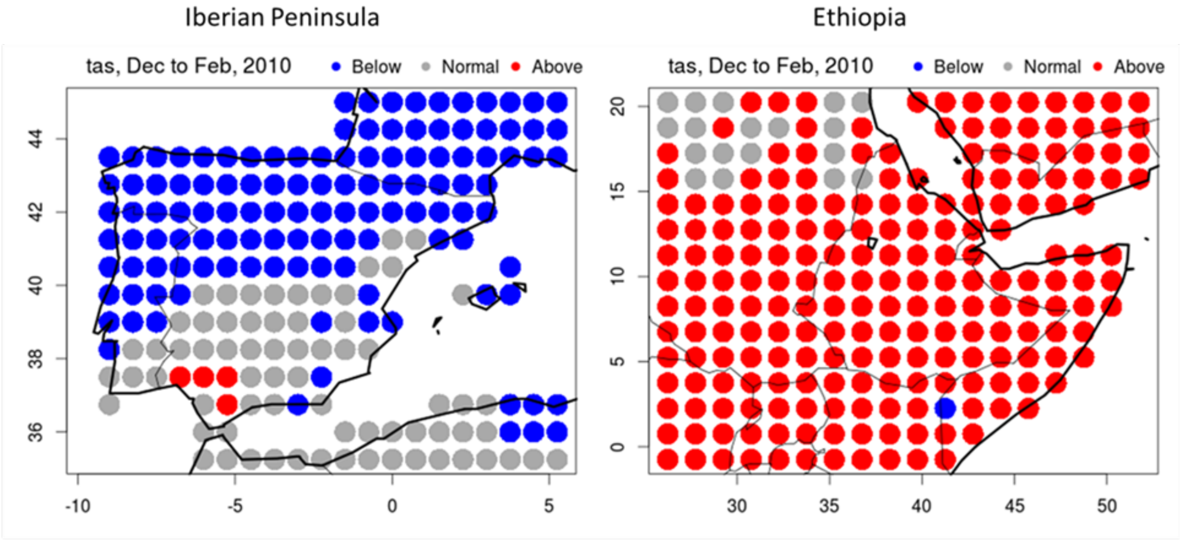
Figure 2a-b Bar graphs showing the predicted likelihood of temperatures being below average, average, and above average for **(a)** The Iberian Peninsula and **(b)** Ethiopia. Skill score (ROCSS) is represented numerically for each tercile. Where skill is below 0 a warning is provided by the use of red lettering.

3.4.3 Bubble plots

In our user needs survey maps emerged as one of the most highly favoured formats for presenting probabilistic information. However, the example presented in this survey represented only the predicted likelihood of the highest tercile, and did not include information about skill, or the likelihood of other terciles. The question of how this information could be effectively integrated into a map therefore remained open. Various possibilities were discussed, including the use of transparency to illustrate skill, and the presentation of separate maps for likelihood and skill (see Lowe et al. (2013) for an example). However, it was ultimately felt that presenting separate skill and probability maps may cause confusion and make it difficult for users to integrate this information. A ‘Bubble Plot’ format that would enable these key pieces of information to be presented on the same map was therefore selected. This map, closely based on visualisations developed by Slingsby et al. (2009) and Jupp et al. (2012), is flexible as it allows different levels of information to be displayed using the same format. In its simplest iteration the map displays only the most likely tercile for each region using a coloured Bubble (a). Information about the predicted likelihood of the most likely tercile may then be added using the size of the Bubble (greater size=greater probability) (b). Alternatively, the predicted likelihood of all three terciles may be presented as a pie chart for each region (c). Finally, shading can be used to illustrate level of skill (ROCSS) for either most likely tercile (d) or all terciles (e), with a darker shade denoting greater skill. To limit the amount of extraneous detail presented, and reduce visual complexity, one can also opt to display information for a small subset of regions only (f).

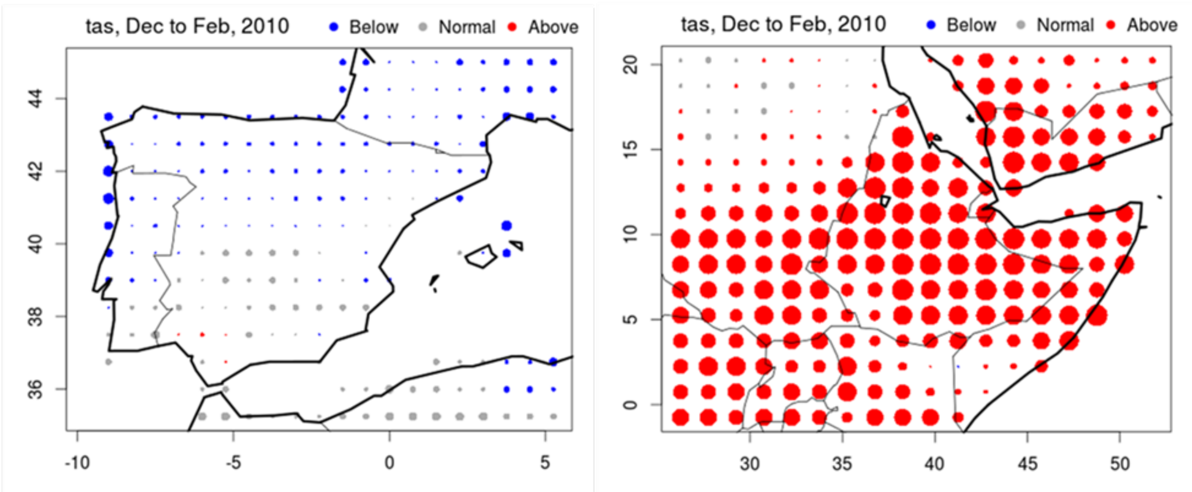
The question of which of these is to be preferred is likely to be contingent on the information needs and preferences of individual users. Version (e) represents the most comprehensive representation, but is also the most visually complex, and thus likely the most difficult to interpret. Although where only a few specific regions are of interest, Version (f) may make this style of plot easier to interpret. Where users are interested in most likely tercile only, Version (d) may be more easily interpreted, but does not illustrate whether a particular tercile is dominant or not. Conversely, Version (c) provides information about the likelihood of each tercile and allows the user to see whether any one tercile dominates, but does not contain information about skill. In regions where skill is negative for a particular time period this could be misleading, and lead to the forecast being used in situations where it would be more appropriate to use climatology (compare for instance Versions (c) and (d) of the plot for the Iberian Peninsula). One way of addressing the problem of negative skill in Versions (a-c) of the plot is to leave blank any region where skill is negative. This does not of course tell users how ‘good’ positive skill is, but it does show where the forecast outperforms climatology, and where climatology should be used instead. Again, it is to be expected that users will differ in terms of how much detail is desired.

In the Decision Lab it is planned that both (c) and (d) will be presented to participants to explore 1) how well these formats are understood; 2) which of them user’s prefer; and 3) how they could be used in decision making. However, the R code for all versions is available.



(a) Most likely tercile only

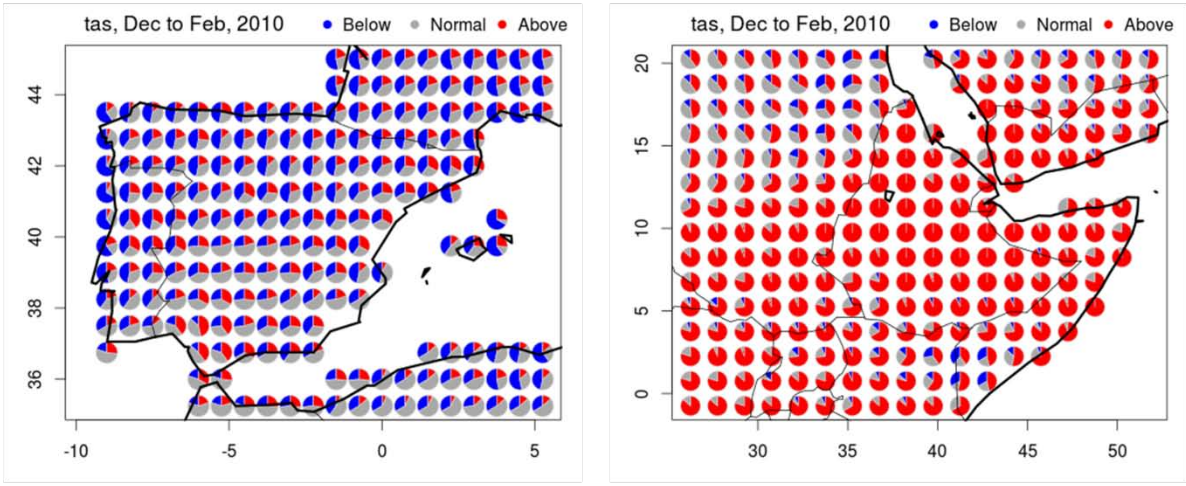
The colour of the bubble shows the most likely tercile for each location (**blue**=below average temperature, **grey**=average temperature, **red**=above average temperature).



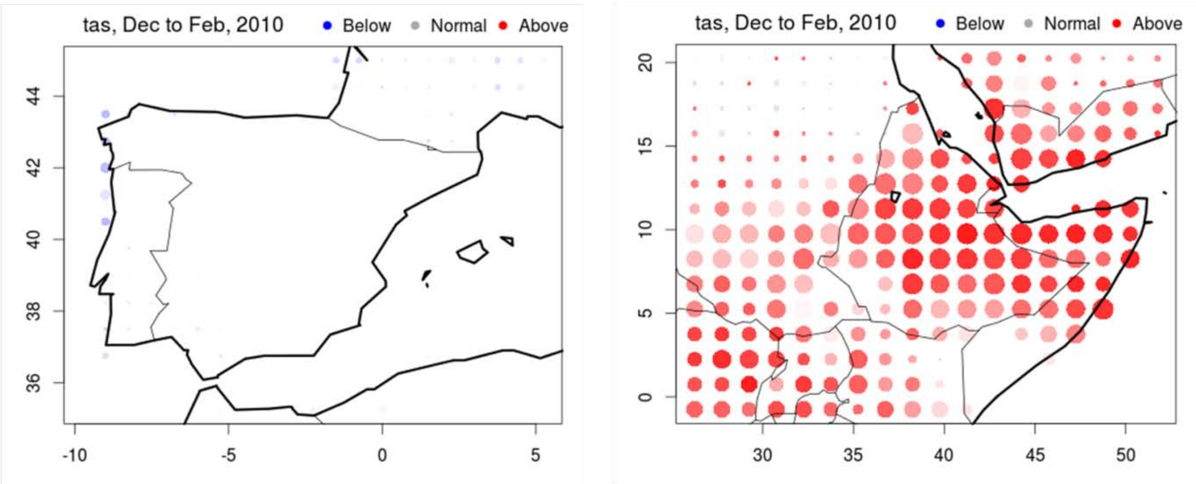
(b) Likelihood of most likely tercile

The size of the Bubble represents the predicted likelihood of most likely tercile for each location. **Larger bubble=greater likelihood**.

Iberian Peninsula   Ethiopia

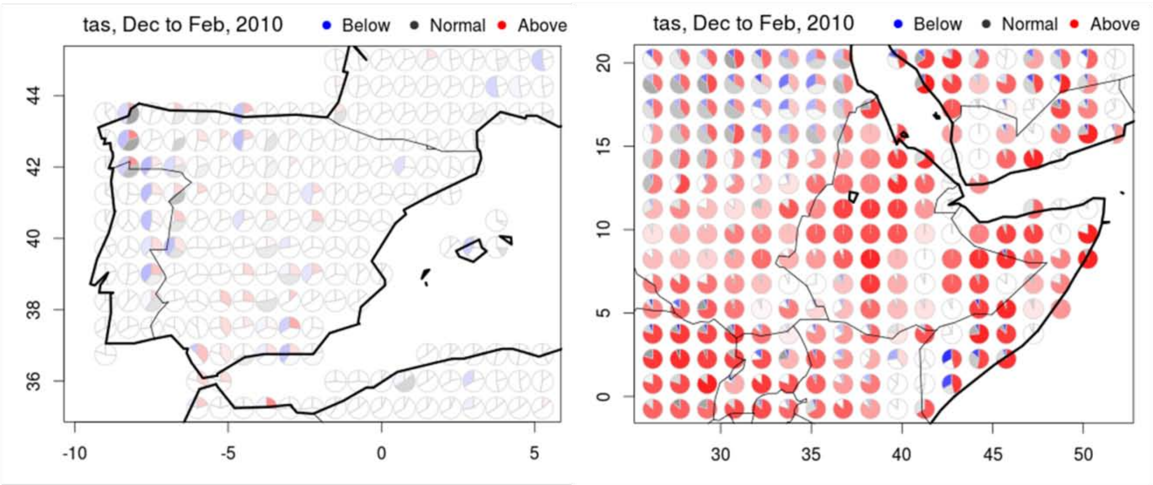


(c) Likelihood of each tercile for each region
Pie charts represent the predicted likelihood of above average (red), average (grey), and below average (blue) temperatures.

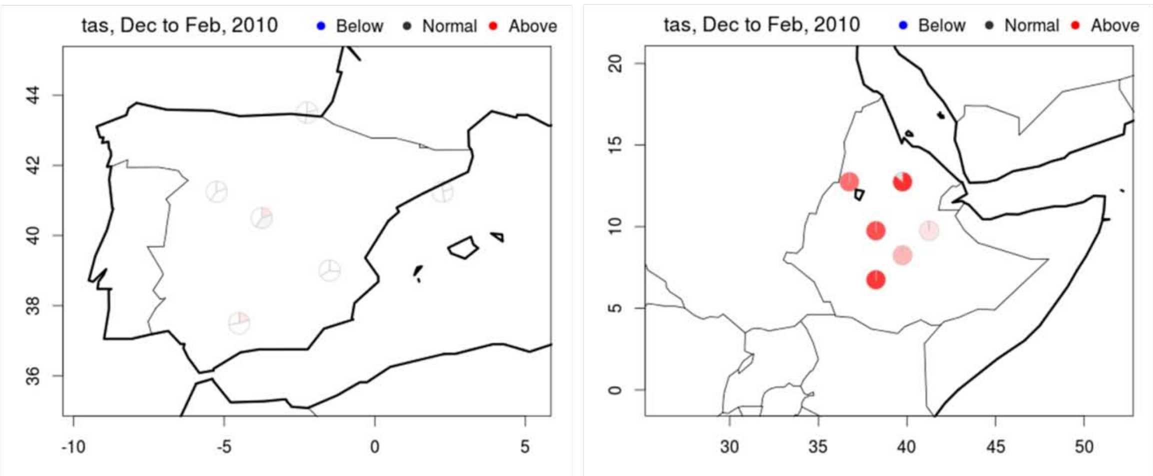


(d) Likelihood of most likely tercile with skill
The shade of the Bubble denotes the skill of the prediction for each region **(darker colour=greater skill).**
Blank regions = no useful information

Iberian Peninsula   Ethiopia



(e) Likelihood of each tercile for each region with skill
Pie charts represent the predicted likelihood of above average (red), average (grey), and below average (blue) temperatures. **The shade of each segment of the pie charts denotes skill for that tercile in that region (darker colour = greater skill).**



(f) Likelihood and skill for each tercile in a subset of regions
The plot shown in (e) is amended to show predictions for a subset of regions only. For the Iberian Peninsula (left) these are : San Sebastian – Igueldo, Barcelona Aeropuerto, Salamanca Aeropuerto, Navacerrada, Albacete Los Llanos, and Cordoba Aeropuerto. For Ethiopia (right) these are: Addis Ababa, Adama, Gondar, Mekele, Awassa, and Dire Dawa.

Figure 3a-f Bubble plots depicting **(a)** most likely tercile only (above average=red, average=grey, below average=blue); **(b)** probability of most likely tercile (larger bubble=greater probability); **(c)** probability of each tercile for each region; **(d)** probability of most likely tercile with skill (darker colour=greater skill); **(e)** probability of each tercile for each region with skill; and **(f)** likelihood of each tercile for a selected subsample of regions of interest (compare to Figure 9). Plots are presented for the Iberian Peninsula (right) Ethiopia (left).

3.4.4 Visualising spread

As previously noted, in our earlier user needs survey a preference for representations showing forecast 'spread' was found amongst those reporting themselves to be more comfortable with complex statistical information. While fan graphs were particularly popular with survey respondents, we took the decision not to include this visualisation in our set for development. This was due to fears that the temporally continuous nature of the visualisation could lead some users to believe that the temporal resolution of the forecast is much higher than it is. Three types of plot representing forecast spread for discrete time periods were therefore created, with each being overlaid on climatology. The first of these visualisations, is a traditional boxplot (Figure 4), with the 'box' around the median representing the interquartile range (middle 50% of ensemble members), and the 'whiskers' representing the extent of the forecast spread (minus outliers). The second is a violin plot (Figure 5), which provides a full pdf, and – while more complex – has the advantage of reflecting the overall shape of the distribution and highlighting any multimodalities in the data. The third is a dot plot (Figure 6), where single dots are used to reflect the mean for each ensemble member for each time period. These are colour coded to reflect the tercile that each falls into. A fourth permutation, where dots representing ensemble members are overlaid onto a violin plot, was also created (Figure 7).

For all four visualisations climatology is represented using greyscale shading, with the middle tercile represented by the darkest shade, and the upper and lower terciles using a lighter shade. To avoid over-interpretation of daily peaks, the daily data has been smoothed by means of a (centred) moving average of 31 days. Therefore, at the location of the plots for each time period, the background shows the monthly mean forecast (the terciles and extremes being computed over members and years). The plots show the spread of the monthly mean forecast (for the different members).

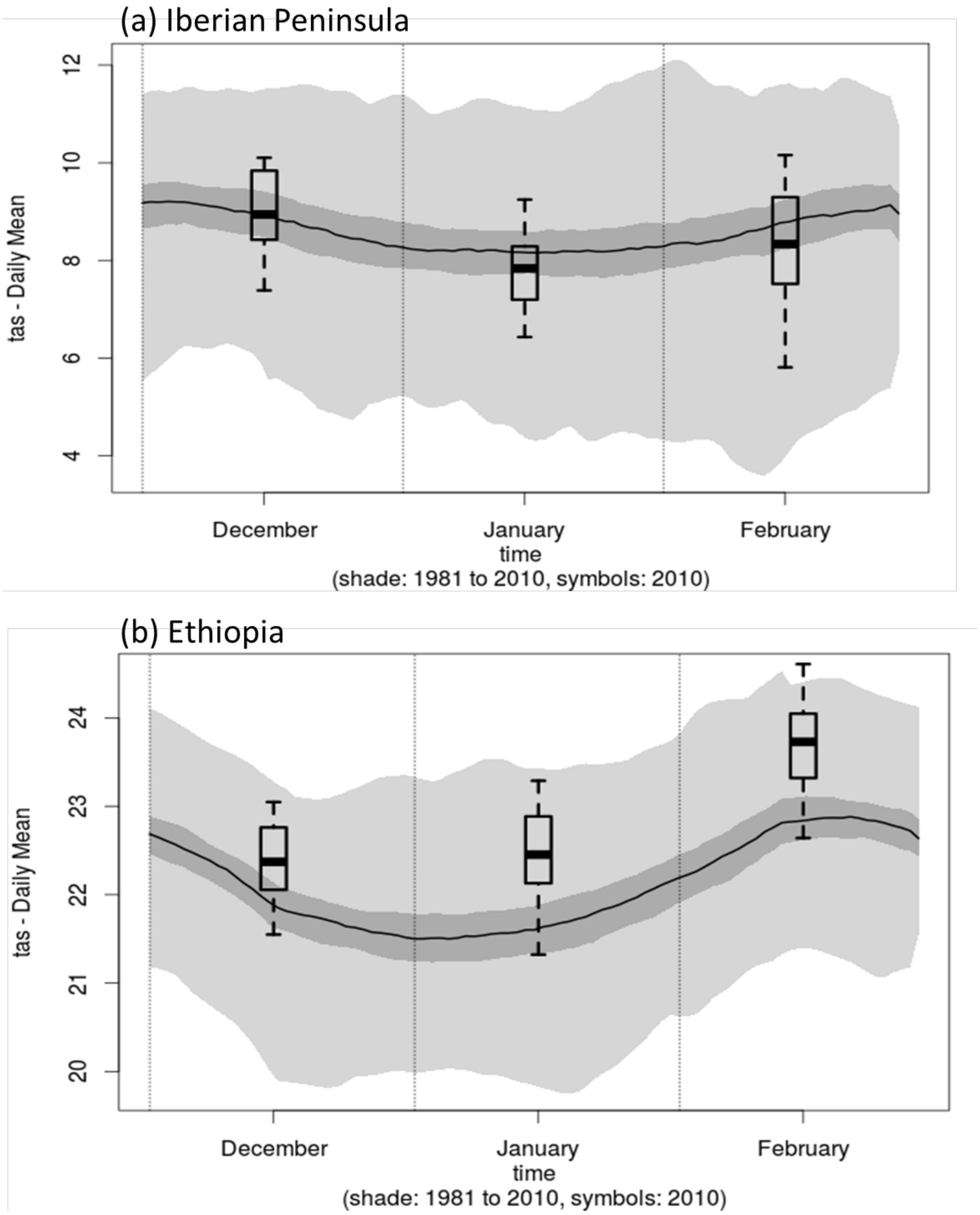


Figure 4 *Boxplot overlaid on climatology for Iberian Peninsula (a) and Ethiopia (b)*. The box around the median represents the interquartile range (middle 50% of ensemble members), while the whiskers represent the extent of the forecast spread (minus outliers). Climatology is represented using greyscale shading, with the middle tercile represented by the darkest shade, and the upper (warmer than average) and lower (cooler than average) terciles by a lighter shade.

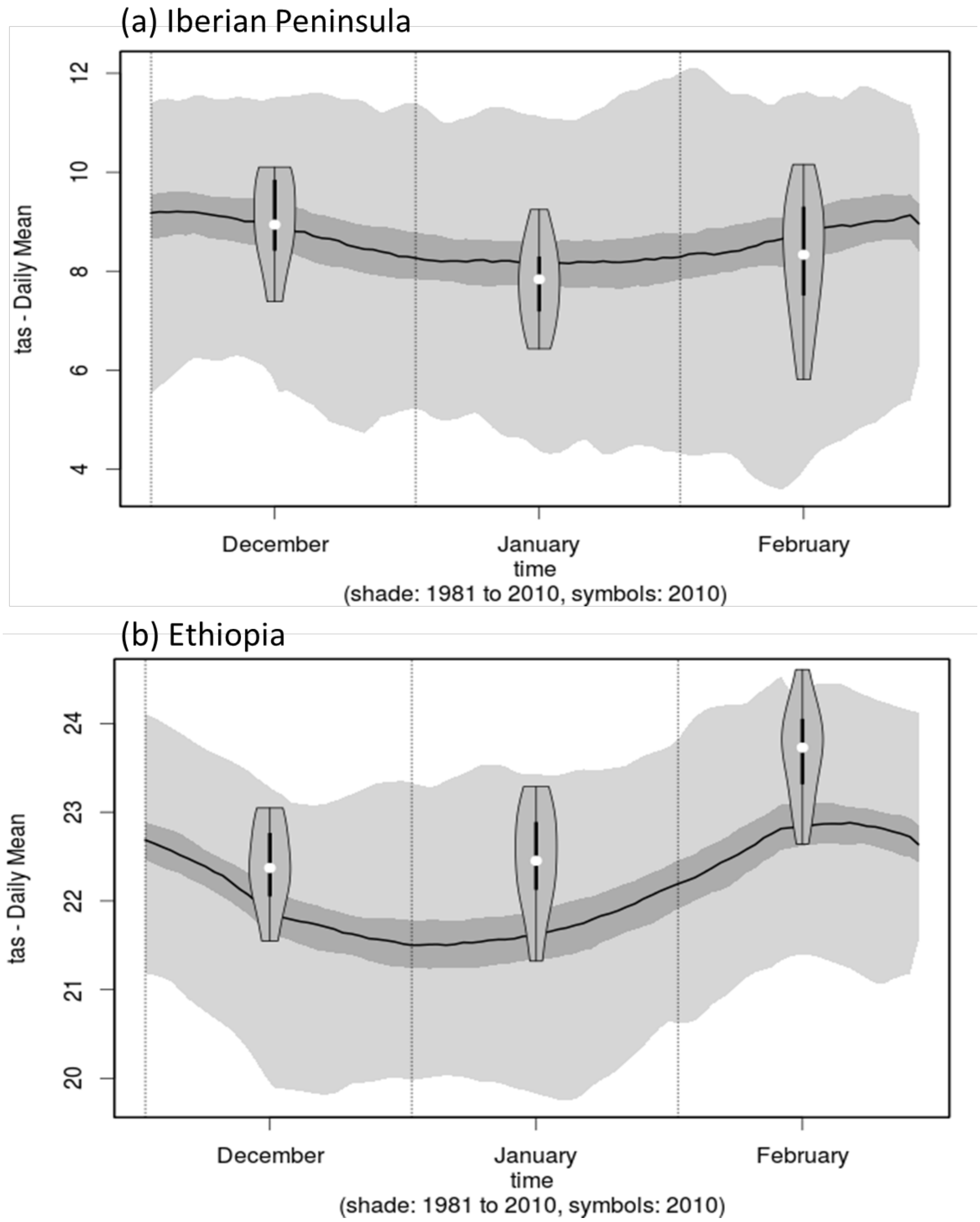


Figure 5 *Violin plot overlaid on climatology for (a) Iberian Peninsula and (b) Ethiopia.* The violin plot represents a full pdf. Climatology is represented using greyscale shading, with the middle tercile represented by the darkest shade, and the upper (warmer than average) and lower (cooler than average) terciles by a lighter shade.

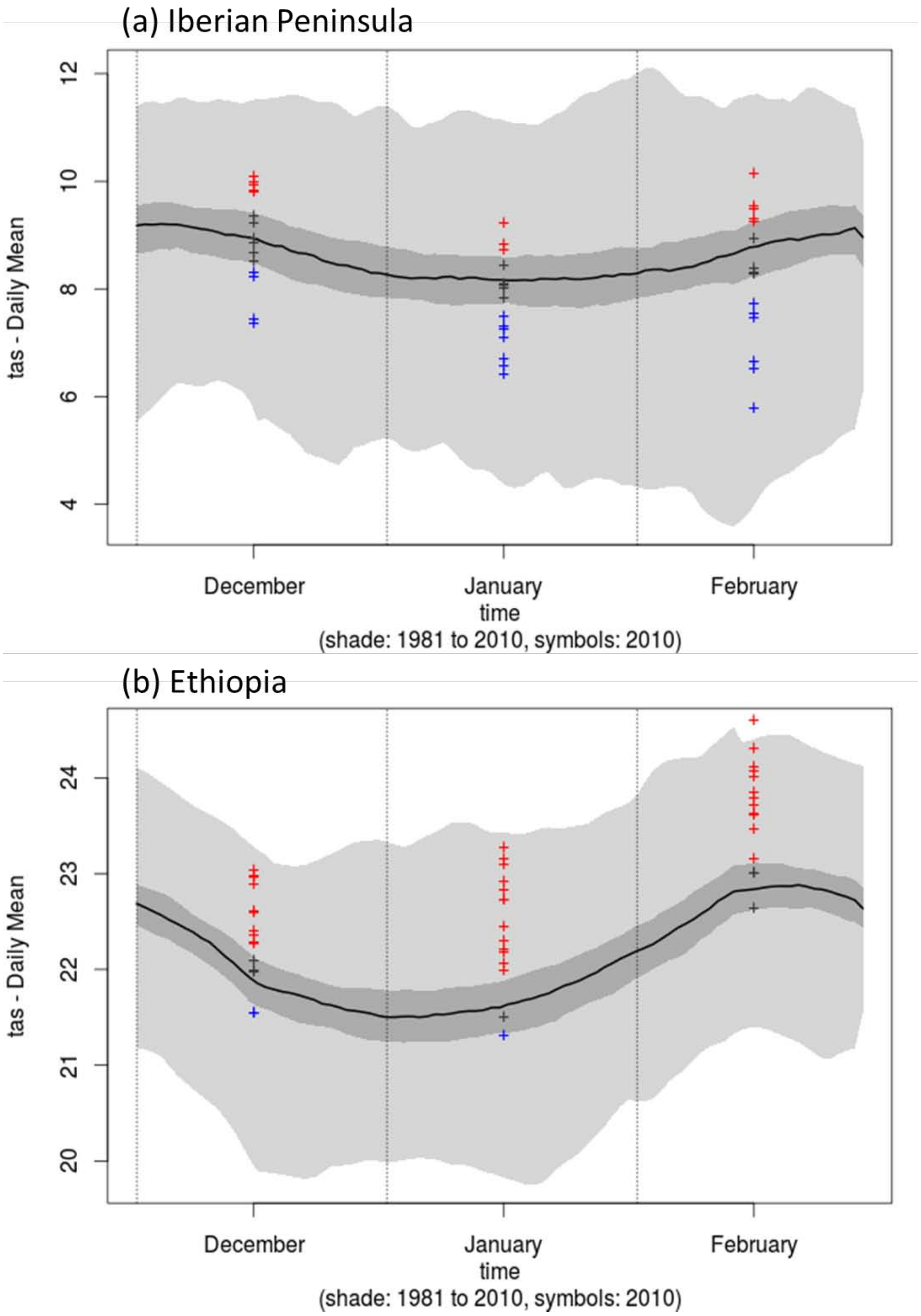


Figure 6 *Dot plot overlaid on climatology for (a) Iberian Peninsula and (b) Ethiopia*. Single dots represent the mean for each ensemble member for each time period. Dots are colour coded to indicate whether they fall into the upper (red), middle (black) or lower (blue) tercile. Climatology is represented using greyscale shading, with the middle tercile represented by the darkest shade, and the upper (warmer than average) and lower (cooler than average) terciles by a lighter shade.

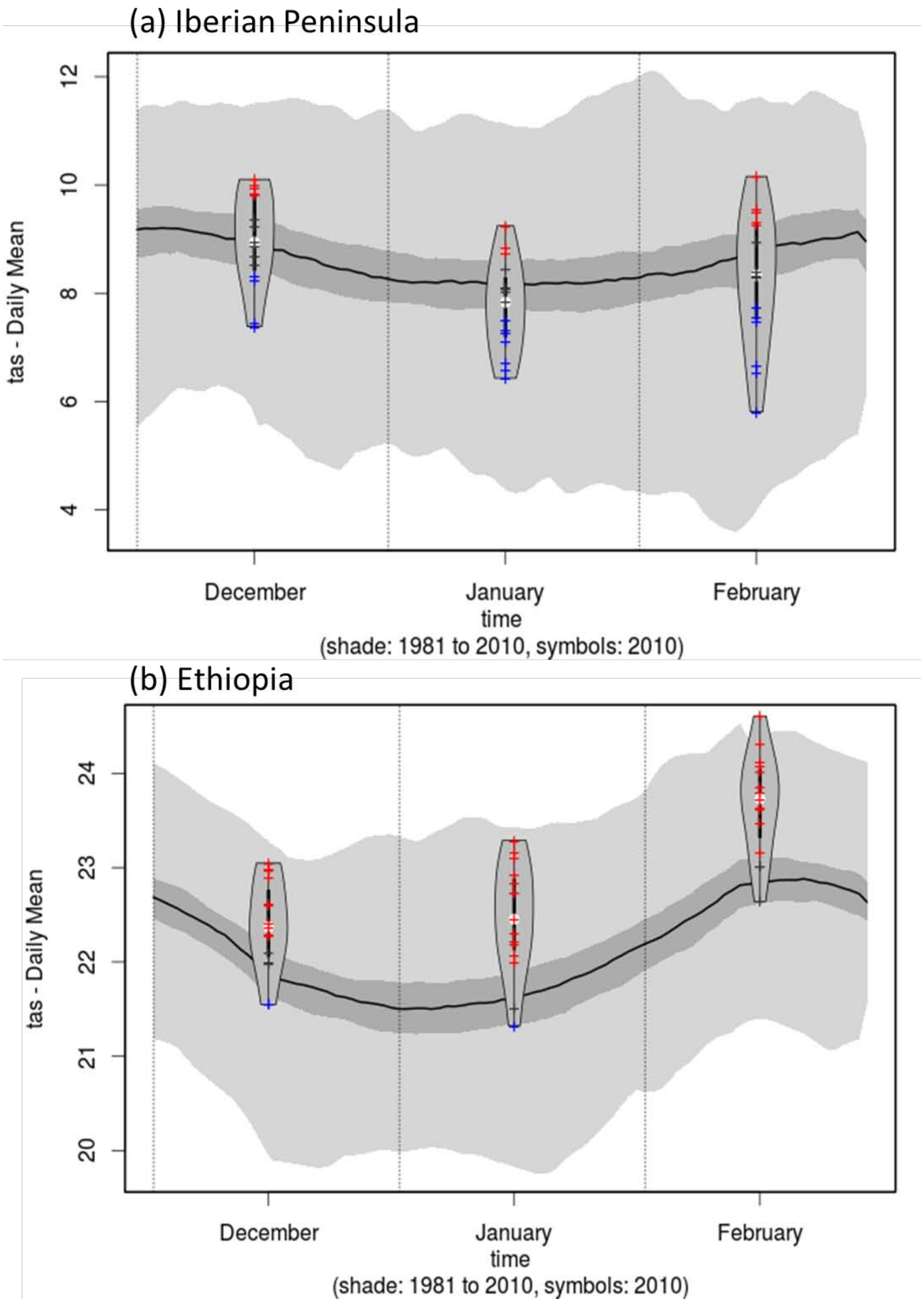


Figure 7 *Dot plot (Figure 6) overlaid on violin plot (Figure 7) for (a) Iberian Peninsula and (b) Ethiopia.*

While the boxplot has the advantage of being a familiar format, for the Decision Lab we have chosen to present the violin plot overlaid with the dot plot (Figure 7). While this is the most complex permutation of the set produced, it has the benefit of visually displaying both overall distribution and tercile. This format was also recommended by one expert end-user within WP33, who reported using the violin plots to communicate with other expert users in their sector. Hence, while this visualisation is unlikely to be suitable for communicating with non-expert end-users, by testing it in the Decision Lab we will assess its usefulness to experts in a range of sectors.

As previously noted, where visualisations show forecast spread rather than discrete categories, presenting skill scores (ROCSS) for each tercile may make it difficult for users to integrate this information. However, where a single skill score for the forecasting system overall (RPSS) is presented it can obscure the fact that predictions may be more skilful for some terciles than for others. Hence, user understanding and preferences with respect to the presentation of skill scores for this type of visualisation requires testing.

3.4.5 Tercile Table

This table, based on that used by Lowe et al. (2013) to present Dengue forecasts, represents an alternative method of presenting information about forecasts for different geographical locations. Here, predicted likelihood of upper, middle and lower terciles is presented as a percentage. Skill is also presented as a numeric value. Two versions of the Table are presented below. The table depicted in Figure 8 uses a single skill score for the whole model (RPSS). The table depicted in Figure 9 by contrast provides skill scores for each tercile (ROCSS).

This format was selected as it offers a way to present information numerically, complementing the visualisations detailed above. Indeed, in our earlier user-needs survey, several respondents indicated a preference for receiving tables in addition to graphs and maps. The information displayed in Figure 9 is the same as that represented in the version of the Bubble Plot illustrated in Figure 3(f). Hence, the two formats could be presented together to enable advanced users to cross reference the visualisation with the corresponding numeric data.

(a) Iberian Peninsula

December-February 2010

City	Cooler than average	Average	Warmer than average	Skill (RPSS)
San Sebastian - Igueldo	47%	33%	20%	-0.213
Barcelona Aeropuerto	53%	27%	20%	-0.143
Salamanca Aeropuerto	40%	40%	20%	-0.174
Navacerrada	40%	40%	20%	-0.180
Albacete Los Llanos	33%	40%	27%	-0.118
Cordoba Aeropuerto	27%	53%	20%	-0.043

(b) Ethiopia

December-February 2010

City	Cooler than average	Average	Warmer than average	Skill (RPSS)
Addis Ababa	0%	0%	100%	0.373
Adama	0%	0%	100%	0.480
Gondar	0%	0%	100%	0.232
Mekele	0%	7%	93%	0.308
Awassa	0%	0%	100%	0.512
Dire Dawa	0%	0%	100%	0.288

Figure 8a-b Tables detailing the predicted likelihood of lower (cooler than average temperature), middle (average temperature) and upper (warmer than average temperature) tercile for a selection of locations in the Iberian Peninsula (a) and Ethiopia (b). A single skill score for the whole model (RPSS) is provided.

(a) Iberian Peninsula

December-February 2010

City	Likelihood			Skill (ROCSS)		
	Cool	Average	Warm	Cool	Average	Warm
San Sebastian - Igueldo	47%	33%	20%	-0.168	-0.411	-0.253
Barcelona Aeropuerto	53%	27%	20%	-0.037	-0.389	0.026
Salamanca Aeropuerto	40%	40%	20%	-0.332	-0.117	0.042
Navacerrada	40%	40%	20%	-0.221	-0.167	-0.063
Albacete Los Llanos	33%	40%	27%	0.016	0.189	-0.068
Cordoba Aeropuerto	27%	53%	20%	0.100	0.011	0.142

(b) Ethiopia

December-February 2010

City	Likelihood			Skill (ROCSS)		
	Cool	Average	Warm	Cool	Average	Warm
Addis Ababa	0%	0%	100%	0.816	0.283	0.663
Adama	0%	0%	100%	0.932	0.522	0.742
Gondar	0%	0%	100%	0.421	0.372	0.695
Mekele	0%	7%	93%	0.547	0.100	0.742
Awassa	0%	0%	100%	0.863	0.511	0.847
Dire Dawa	0%	0%	100%	0.426	0.139	0.837

Figure 9a-b Tables detailing the predicted likelihood of lower (cool), middle (average) and upper (warm) tercile for a selection of locations in the Iberian Peninsula (a) and Ethiopia (b). Skill scores for each tercile (ROCSS) are provided.

3.4.6 Simple formats using evaluative categories and text

With the possible exception of the bar graph and the simpler variants of the Bubble plot, the communication strategies outline above are not likely to be readily understood by users without experience in using statistical information. In the absence of an existing set of empirically supported guidelines for communicating about both forecast probability and skill with this group, we are therefore faced with the challenge of identifying strategies that can fulfil this goal. Through discussion amongst work package partners and a review of the literature on risk communication, possible methods of integrating information about likelihood and skill for this group included the use of evaluative categories (where, for example, numeric information is coded as 'High', 'Medium', 'Low'), simple tables, and verbal descriptions. Three pilot formats aimed at those less comfortable with using statistical information have therefore been proposed. However, as this represents 'new ground', it is our intention to pilot test these formats with public samples in the UK and France prior to the Decision Lab with stakeholders. This will allow us to 1) select the most promising of these for inclusion in the Decision Lab; and 2) make adjustments based on the outcome of these preliminary trials.

(a) Confidence Index incorporating probability and skill

Forecast

"Temperatures will be over 25°C on average next june"

Confidence index

Or 2

Confidence index scale

		Skill		
		Low	**Medium**	**High**
Likelihood	**Low (30-40%)**	No signal	No signal	
	Medium (40-50%)	No signal		
	High (>50%)			

		Low	**Medium**	**High**
Likelihood	**Low (30-40%)**	No signal 0	No signal 1	2
	Medium (40-50%)	No signal 1	2	3
	High (>50%)	2	3	4

Figure 10 Forecast for temperature threshold exceedance presented using a colour-coded Confidence-Index incorporating probability and skill. Likelihood is classified as Low, Medium or High and weighted by skill (Low, Medium, High) to provide a Confidence Index. Confidence is rated on a scale of 0-5 where 0 indicates Low skill and Low probability of exceedance, while 5 would represent a hypothetical perfect forecast. Where Confidence is lower than 2 a 'No Signal' message is given.

This format is based on the Confidence Index used by MeteoFrance for their weather forecasts, weights probability by skill to produce a score and associated colour code, to signal how 'confident' one may be that a threshold of interest will be exceeded. In this example the likelihood of mean temperature exceeding a particular value is coded as Low

(30-40%), Medium (40-50%) or High (over 50%). Skill is also coded as Low (> 0 and < 0.20), Medium (≥ 0.20 and < 0.50) or High (≥ 0.50). Here a Confidence Score of 5 would indicate a hypothetical perfect forecast, while one of 0 would indicate Low likelihood and Low skill. For Confidence scores of 0 or 1 a 'No Signal' message is given, whilst 2 (amber), 3 (yellow), and 4 (green) indicate progressively higher confidence that the threshold will be exceeded. In keeping with recommendations made in the environmental risk communication literature (Kloprogge et al., 2007), this format is designed to progressively disclose detail about the forecast. The 'topline' consists of just the Confidence score and its corresponding colour code (or a 'no signal' message). This is followed by scales showing how the score is derived. The goal of this format is to provide information about uncertainty and confidence without presenting detailed information about probability and skill.

(b) Simple table with skill category

The table below shows how likely it is that [SEASON] temperatures for [REGION] will be warmer than average, average, or cooler than average, according to our forecast. As it is more difficult to predict what the temperature will be like for some month's than for others, the performance of the forecast can vary across the year. However, by looking at how well the forecast has performed in the past we can rate how much confidence we can have in it.

Temperature	Likelihood	Confidence
Warmer than average	60%	Low
Average	20%	
Cooler than average	20%	

For [season] our forecast has a **Low** Confidence rating. but the actual likelihood of temperatures being above average, average or below average is very often different than predicted.

Confidence Scale

No Confidence The forecast doesn't offer any useful information at this time

Low Confidence The forecast is better than chance, but the actual likelihood of temperatures being above average, average or below average is **very often** different than predicted.

Med Confidence The forecast is better than chance, but the actual likelihood of temperatures being above average, average or below average is **often** different than predicted.

High Confidence The forecast is better than chance, but the actual likelihood of temperatures being above average, average or below average is **sometimes** different than predicted.

Figure 11 Simple table with skill category. This table details the forecast probability of it being warmer than average, average, or cooler than average, with a categorical rating of skill (RPSS), referred to as Confidence, on a scale of No skill/confidence (RPSS ≤ 0), Low skill/confidence (RPSS > 0 and < 0.20), Medium skill/confidence (RPSS ≥ 0.20 and < 0.50) and High skill/confidence (RPSS ≥ 0.50).

(c) Text description with skill categories

Our forecast predicts that for [TIME PERIOD] it is very likely (95% chance) that average temperatures will fall between 10^{o}C and 15^{o}C. However it should be kept in mind that forecasts are not perfectly reliable, and perform better for some seasons than others. By looking at how well forecasts have predicted temperatures for previous years, we can measure how they are performing and give them a Confidence score. For [Season] Confidence scores are **Low**. This means that while forecasts for [Season] tend to provide a better guide to future temperature than just looking at historical averages, they often make errors.

Confidence Score

No Confidence	The forecast doesn't offer any useful information at this time
Low Confidence	The forecast is better than chance, but the actual likelihood of temperatures being above average, average or below average is **very often** different than predicted.
Med Confidence	The forecast is better than chance, but the actual likelihood of temperatures being above average, average or below average is **often** different than predicted.
High Confidence	The forecast is better than chance, but the actual likelihood of temperatures being above average, average or below average is **sometimes** different than predicted.

Figure 12 Text description of forecast with skill represented as a Confidence Rating: No Confidence (RPSS ≤ 0), Low Confidence (RPSS > 0 and < 0.20), Medium Confidence (RPSS ≥ 0.20 and < 0.50) and High Confidence (RPSS ≥ 0.50).The forecast is presented by describing the range of temperatures falling within a 95CI of the forecast.

Here predicted temperature is presented as a 95CI confidence range, with 'Confidence' (skill) being classified as None, Low, Medium and High. The phrasing used is based on that developed by the IPCC for classifying likelihood versus amount and quality of information (Mastrandrea et al., 2010). However, as one criticism of the IPCC scheme has been that phrases such as 'very likely' are open to a variety of interpretations, we have followed Budescu et al. (2009) recommendation that verbal descriptions of likelihood be accompanied by numeric details.

Summary: Strategies for communicating levels of confidence and uncertainty in S2D predictions

- Based on the findings of our user needs survey findings and discussion amongst work package partners we identified and developed the following formats for users with moderate to high comfort with statistics:
  - **3.4.1 Tercile Plot**: Showing explicitly how past predictions have corresponded observations (skill represented using ROCSS).
  - **3.4.2 Tercile Bar graph**: A simpler representation of likelihood of each tercile with corresponding skill score (ROCSS)
  - **3.4.3 Bubble Plot (Map)**: Map illustrating forecast information using a 'bubble' for each region (skill represented using ROCSS). This visualisation allows uncertainty to be represented at different levels of complexity.
  - **3.4.4 Representations of Spread**: Boxplot, Violin plot and Dot plot representing forecast distribution (skill represented using either ROCSS for each tercile, or RPSS for whole forecast).
  - **3.4.5 Table**: Table numerically representing likelihood of each tercile and skill score (skill represented using either ROCSS for each tercile, or RPSS for whole forecast)
- For users who are less familiar with statistical information the following formats have been proposed, and will undergo preliminary testing. In these instances skill is rated using a None (≤ 0), Low (>0 and ≤ 0.20), Medium (≥0.20 and <0.50), High (≥ 0.50) scale of categorisation.
  - **3.4.6a Confidence Index combining likelihood and skill** (skill to be represented using either ROCSS or RPSS)
  - **3.4.6b Simple Table** (skill to be represented using RPSS)
  - **3.4.6c Verbal description** (skill to be represented using RPSS)

3.5 Future Directions: Plans for testing communication formats

The objective of our next Work Package Task (T33.4) is to test the communication strategies detailed in this report with end users. As noted above, prior to the Decision Labs with relevant stakeholders we will pilot the formats described in 3.4.6 with members of the general public in France and the UK. While this will not tell us how useful these formats might be to those using the information operationally, it will test how non-experts understand and interpret the information. Therefore allowing us to a) select that which is best understood for the Decision Lab; and b) make modifications based on feedback. By sampling from France and the UK we will also be able to ascertain whether understanding of skill is greater in France (where a Confidence Index is presented with the weather forecast) than the UK (where weather forecasts are deterministic); potentially highlighting where national differences should be considered. This pilot testing will then be followed by a Decision Lab with relevant stakeholders, where participants in multiple sectors will be presented with the communication strategies selected. The Decision Lab will be built using online survey and experiment software so that it can be run both online and at workshops. Questions presented to participants will examine: a) how well these formats are understood; b) which formats user's prefer; and c) how users would make use of them in their decision making. The findings of the Decision Lab will enable us to make a series of recommendations for good practice in communicating uncertainty in S2D.

3.6 References

BUDESCU, D. V., BROOMELL, S. & POR, H.-H. 2009. Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science,* 20**,** 299-308.

DARON, J. 2014. Visualising climate model output to communicate climate risk information to users across Africa.

GREGORY, R., DIECKMANN, N., PETERS, E., FAILING, L., LONG, G. & TUSLER, M. 2012. Deliberative disjunction: Expert and public understanding of outcome uncertainty. *Risk analysis,* 32**,** 2071-2083.

JOLLIFFE, I. T., STEPHENSON, D. B. 2003. Forecast Verification. A practitioner's guide in Atmospheric Science. Wiley & Sons.

JUPP, T. E., LOWE, R., COELHO, C. A. & STEPHENSON, D. B. 2012. On the visualization, verification and recalibration of ternary probabilistic forecasts. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* 370**,** 1100-1120.

KAIN, D. & COVI, M. 2013. Visualizing complexity and uncertainty about climate change and sea level rise. *Communication Design Quarterly Review,* 1**,** 46-53.

KAYE, N., HARTLEY, A. & HEMMING, D. 2012. Mapping the climate: guidance on appropriate techniques to map climate variables and their uncertainty. *Geoscientific Model Development,* 5**,** 245-256.

KLOPROGGE, P., SLUIJS, J. P. & WARDEKKER, J. A. 2007. *Uncertainty communication: issues and good practice*, Copernicus Institute for Sustainable Development and Innovation, Utrecht University Utrecht.

LORENZ, S., DESSAI, S., FORSTER, P. & PAAVOLA, J. 2015. Tailoring the visual communication of climate projections for local adaptation practitioners in Germany and the UK.

LOWE, R., BAILEY, T. C., STEPHENSON, D. B., JUPP, T. E., GRAHAM, R. J., BARCELLOS, C. & CARVALHO, M. S. 2013. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statistics in medicine,* 32**,** 864-883.

MASTRANDREA, M. D., FIELD, C. B., STOCKER, T. F., EDENHOFER, O., EBI, K. L., FRAME, D. J., HELD, H., KRIEGLER, E., MACH, K. J. & MATSCHOSS, P. R. 2010. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. *Intergovernmental Panel on Climate Change (IPCC)*.

PETERS, E., DIECKMANN, N. F., VÄSTFJÄLL, D., MERTZ, C., SLOVIC, P. & HIBBARD, J. H. 2009. Bringing meaning to numbers: the impact of evaluative categories on decisions. *Journal of experimental psychology: applied,* 15**,** 213.

SLINGSBY, A., LOWE, R., DYKES, J., STEPHENSON, D., WOOD, J. & JUPP, T. A pilot study for the collaborative development of new ways of visualising seasonal climate forecasts. Proc. 17th Annu. Conf. of GIS Research UK, Durham, UK, 1â€“3 April 2009, 2009. Citeseer.

STEPHENS, E. M., EDWARDS, T. L. & DEMERITT, D. 2012. Communicating probabilistic information from climate model ensembles—lessons from numerical weather prediction. *Wiley Interdisciplinary Reviews: Climate Change,* 3**,** 409-426.

TAYLOR, A. L**.** & DESSAI, S. (2014) Deliverable 33.1 Report on survey of end-user needs for improved uncertainty and confidence level information, EUPORIAS **http://euporias.eu/sites/default/files/deliverables/D33.1_Final.pdf**

TAYLOR, A. L., DESSAI, S, BUONTEMPO, C., & DUBOIS, G. (2014) Deliverable 33.2 Report summarising review of existing approaches for communicating confidence and uncertainty, EUPORIAS **http://euporias.eu/sites/default/files/deliverables/D33.2_Final.pdf**

WEEDON, G. P., BALSAMO, G., BELLOUIN, N., GOMES, S., BEST, M. J., & VITERBO, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resources Research, 50(9), 7505-7514.

3.7. Acknowledgements

In addition to the contributions made by partners in Work Package 33 and Work Package 32, we would like to thank Aidan Slingsby (City University), Melanie Davis (IC3) and Rachel Lowe (IC3) for their input and feedback during the process of selecting and developing the strategies for communicating uncertainty presented in this report.

4. Lessons Learnt

5. Links Built

With other deliverables:

- The process of selecting and developing the strategies for communication detailed in this report has been informed by D33.1 and D33.2.

Links with other work packages:

- We have collaborated with Jesus Fernandez and Maria Dolores Frias (WP32, University of Cantabria) to develop and produce the visualisations detailed in this report.

6. Supplementary material: R Code for 3.4.1 – 3.4.5

EUPORIAS Visualisation Plots

Universidad de Cantabria

27/04/2015

Contents

1 Proposed visualizations

This document shows some sample R functions to visualize seasonal forecast information with different levels of complexity.

1.1 Requirements

The functions have been incorporated into the development branch of downscaleR. Get the latest version using the following code. This has to be done **just once** in this order:

```
library(devtools)
install_github("SantanderMetGroup/downscaleR.java@stable")
install_github("mdfrias/downscaleR")
install_github("SantanderMetGroup/ecomsUDG.Raccess@stable")
```

We'll need to load some packages and log into the ECOMS UDG to download the data:

```
library(downscaleR)
```

```
## Loading required package: rJava
## Loading required package: downscaleR.java
## NetCDF Java Library v4.3.22 (27 May 2014) loaded and ready
## Loading required package: maps
## Loading required package: vioplot
## Loading required package: sm
## Package 'sm', version 2.2-5.4: type help(sm) for summary information
## Loading required package: mapplots
## downscaleR version 0.5-2 (22-Jan-2015) is loaded
## WARNING: Your current version of downscaleR (v0.5-2) is not up-to-date
## Get the latest stable version (0.6.0) using <devtools::install_github('SantanderMetGroup/downscaleR@s
```

```
library(ecomsUDG.Raccess)
```

```
## ecomsUDG.Raccess version 2.2-6 (27 Jan 2015) is loaded
```

```
loginECOMS_UDG("username", "password")
```

1.2 Sample data

We can load some sample data from the ECOMS-UDG. E.g. for surface temperature:

```
var <- "tas"
year.ini <- 1982
year.end <- 2010
year.target <- 2010
season <- c(12,1,2)
lead.month <- 1
members <- 1:15
```

The plots shown in this document are focused on Spain:

```
lonlim <- c(-10,5)
latlim <- c(35,45)
```

but you could choose any other region by setting appropriate lat-lon boundaries. For example, for Peru:

```
lonlim <- c(-83,-66)
latlim <- c(-20,0)
```

or Ethiopia

```
lonlim <- c(26,52)
latlim <- c(-1,20)
```

We are ready to load the predictions and observations. We can use `downscaleR` functions to interpolate the data to a common grid.

```
prd <- loadECOMS(dataset = "System4_seasonal_15", var=var,
                 lonLim=lonlim, latLim=latlim, season=season, years=year.ini:year.end,
                 leadMonth=lead.month, members=members, time="DD"
)
```

```
## [2015-04-25 18:36:04] Defining homogeneization parameters for variable "tas"
## [2015-04-25 18:36:09] Defining geo-location parameters
## [2015-04-25 18:36:09] Defining initialization time parameters
## [2015-04-25 18:36:10] Retrieving data subset ...
## [2015-04-25 18:37:28] Done
```

```
obs <- loadECOMS(dataset = "WFDEI", var=var,
                 lonLim=lonlim, latLim=latlim, season=season, years=year.ini:year.end
)
```

```
## [2015-04-25 18:41:03] Defining homogeneization parameters for variable "tas"
## [2015-04-25 18:41:04] Defining geo-location parameters
## [2015-04-25 18:41:04] Defining time selection parameters
## [2015-04-25 18:41:04] Retrieving data subset ...
## [2015-04-25 18:41:21] Done
```

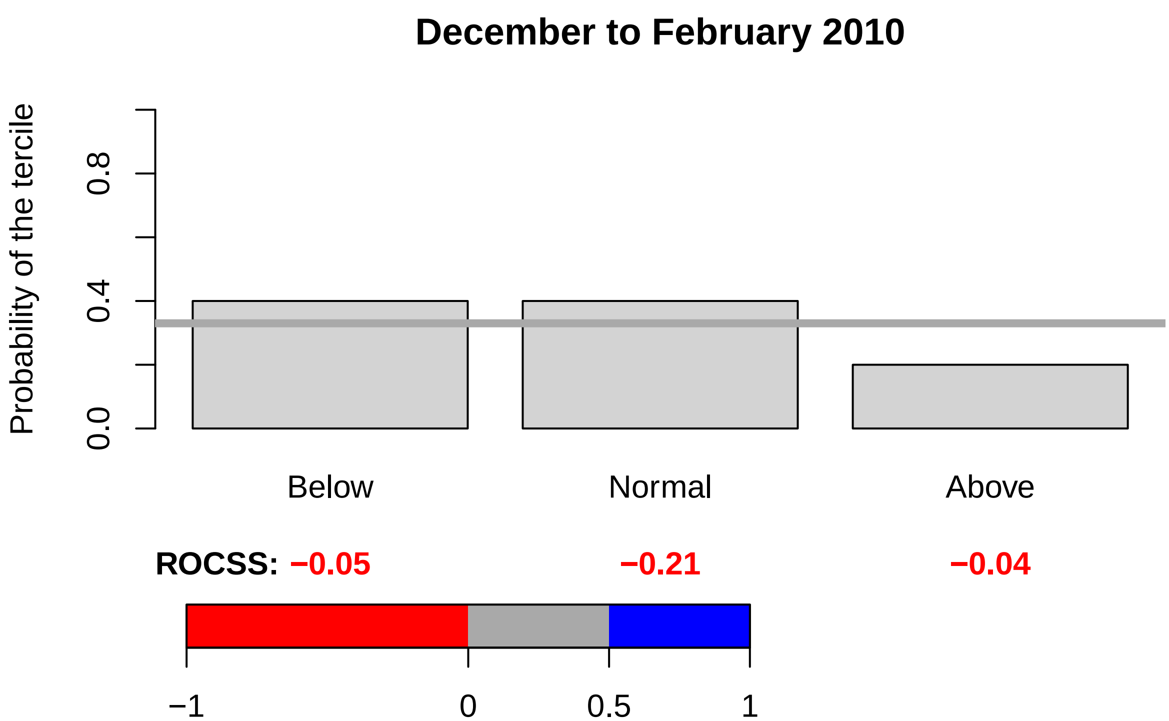
```
# Interpolation to model grid
obs <- interpGridData(obs, new.grid = getGrid(prd), method = "nearest")
```

```
## [2015-03-05 18:42:47] Performing nearest interpolation... may take a while
## [2015-03-05 18:43:05] Done
## Warning messages:
## 1: In interpGridData(obs, new.grid = getGrid(prd), method = "nearest") :
##   The new longitudes are outside the data extent
```

1.3 Visualization functions

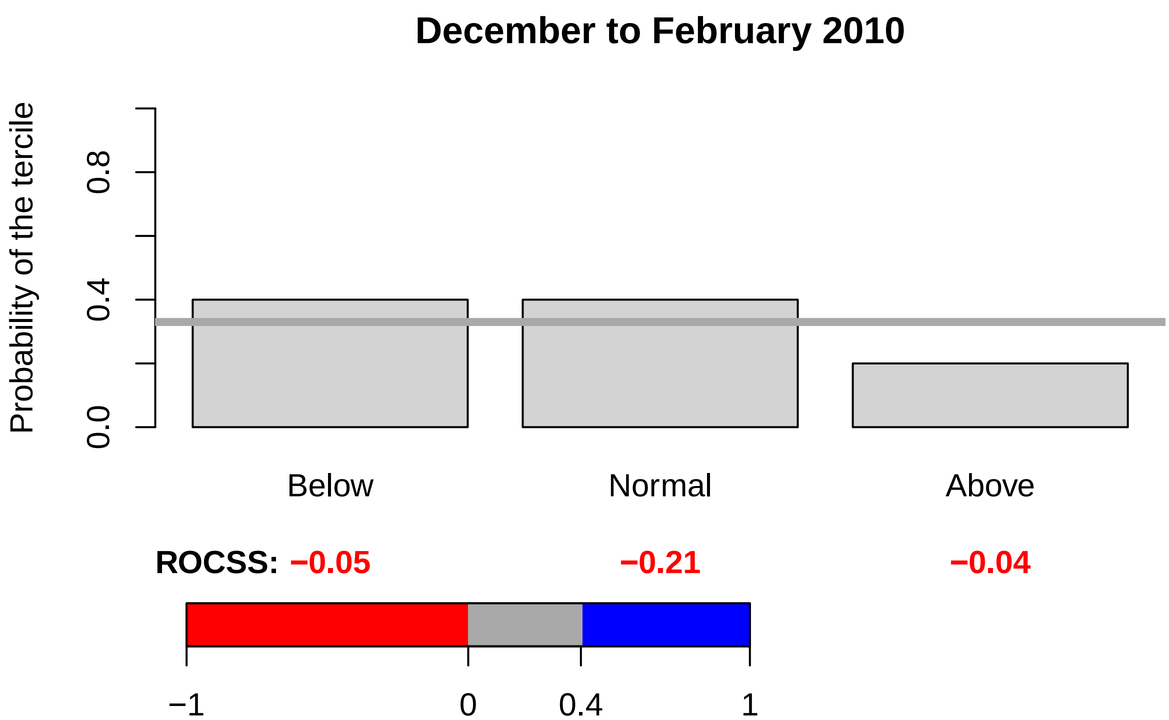
1.3.1 Tercile bar plot

```
tercileBarplot(prd, obs, year.target)
```



Depending on the area, the positive score threshold to be remarked can be different. It is possible to indicate a positive score threshold to highlight in a different color those values of the score above a particular value. 0.5 is the default value.

```
tercileBarplot(prd, obs, year.target, score.threshold=0.4)
```

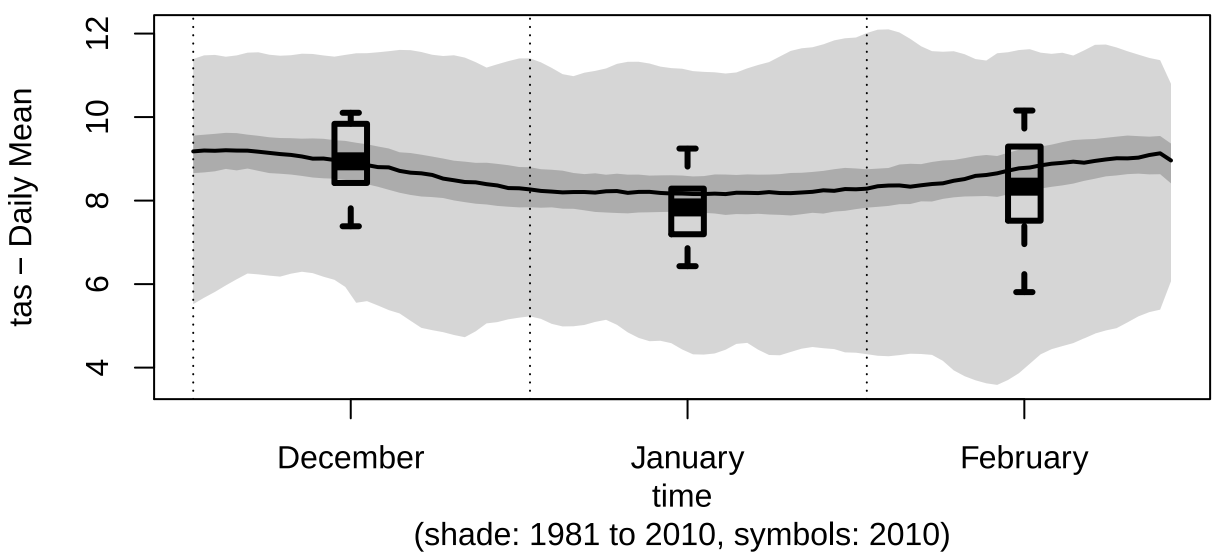


1.3.2 Box plots on climatology fan chart

In this plot the background represents the climatology for the forecast period. The shaded areas show the central tercile (dark shade) and the maximum and minimum (light shade). To avoid overinterpretation of daily peaks, the daily data has been smoothed by means of a (centered) moving average of 31 days. Therefore, at the location of the boxplots, the background shows the monthly mean forecast (the terciles and extremes being computed over members and years).

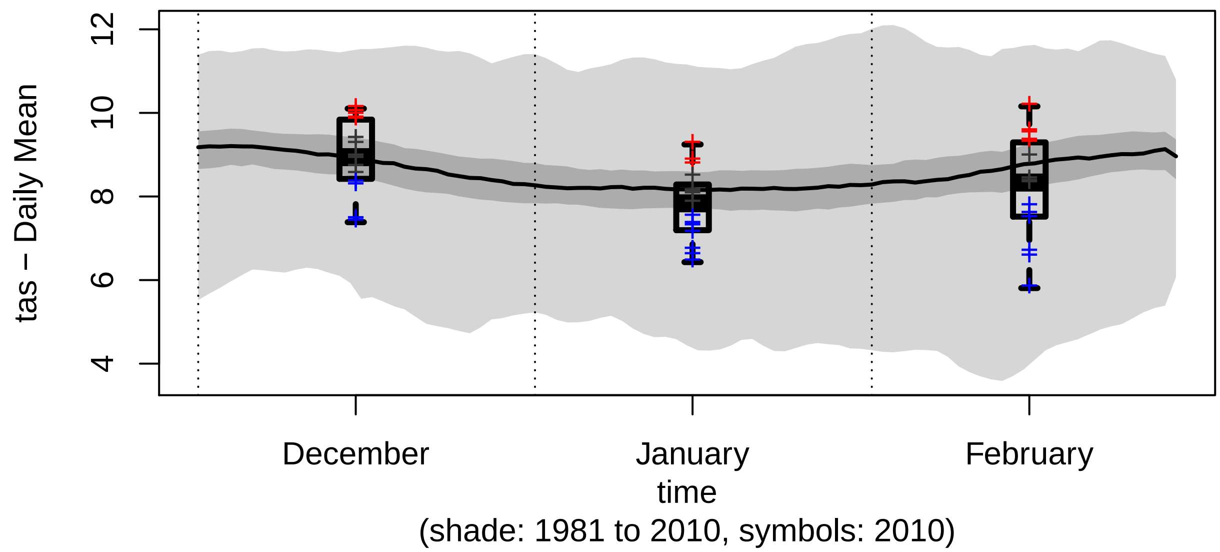
The boxplots show the spread of the monthly mean forecast (for the different members).

```
spreadPlot(prd, obs, year.target)
```

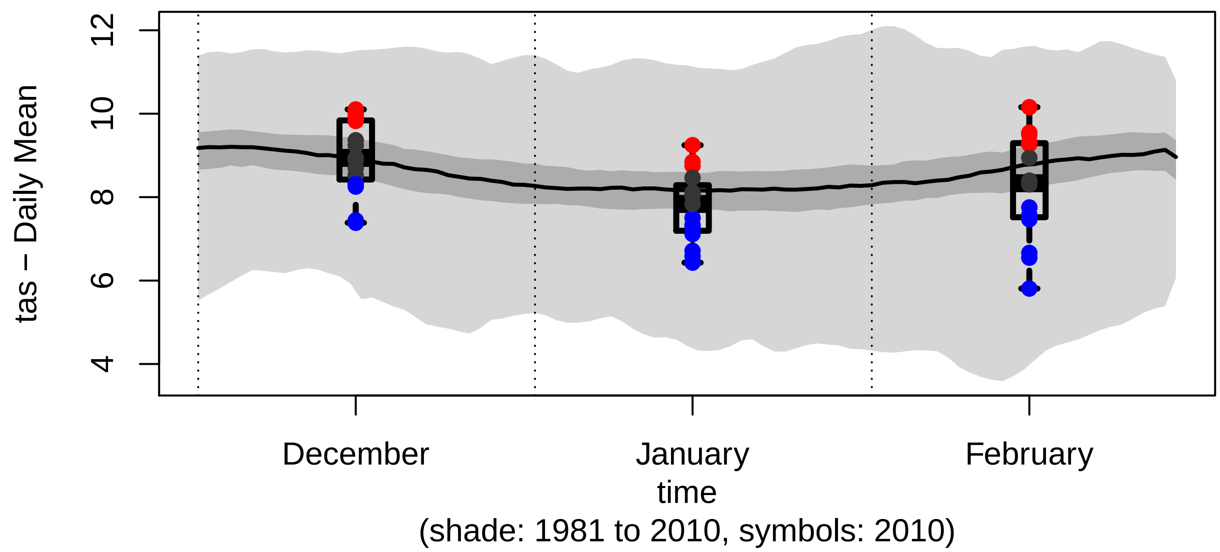


You could also add to the boxplot the values of the ensemble members using crosses or any other symbol (pch option).

```r
spreadPlot(prd, obs, year.target, add.points=T)
```

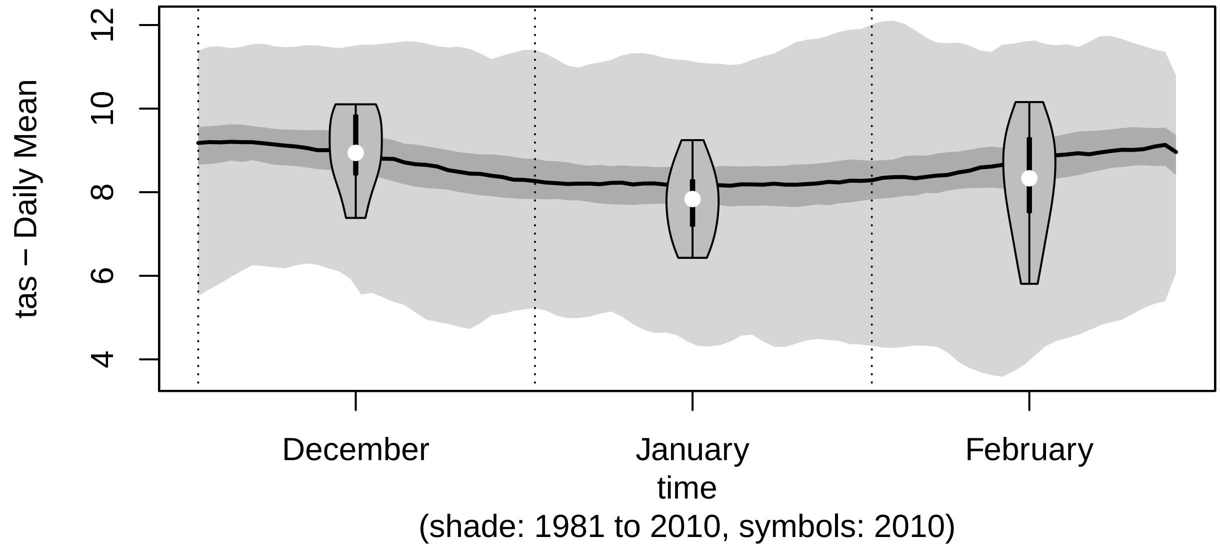



```r
spreadPlot(prd, obs, year.target, add.points=T, pch=19)
```

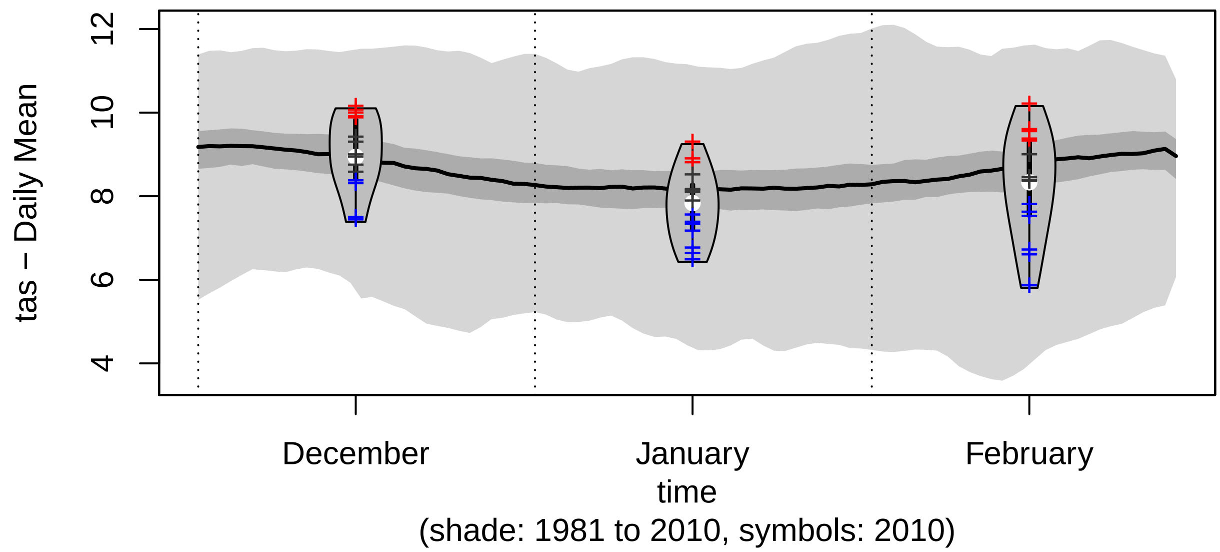


The boxplots can be replaced by *violin plots*, to unveil multimodalities in the data.

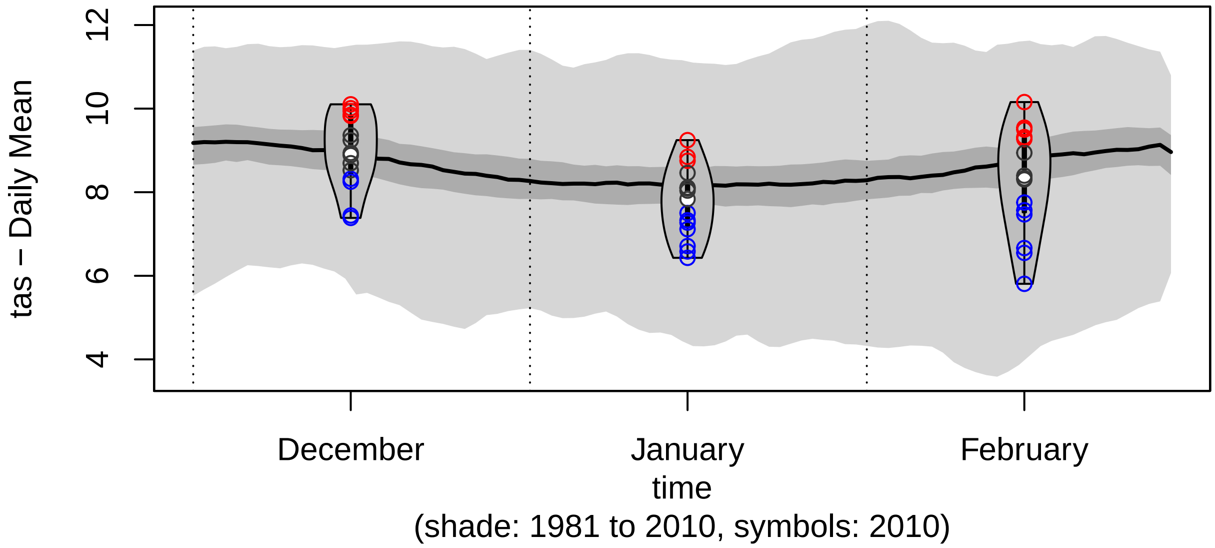
```r
spreadPlot(prd, obs, year.target, violin=T)
```




```r
spreadPlot(prd, obs, year.target, violin=T, add.points=T)
```

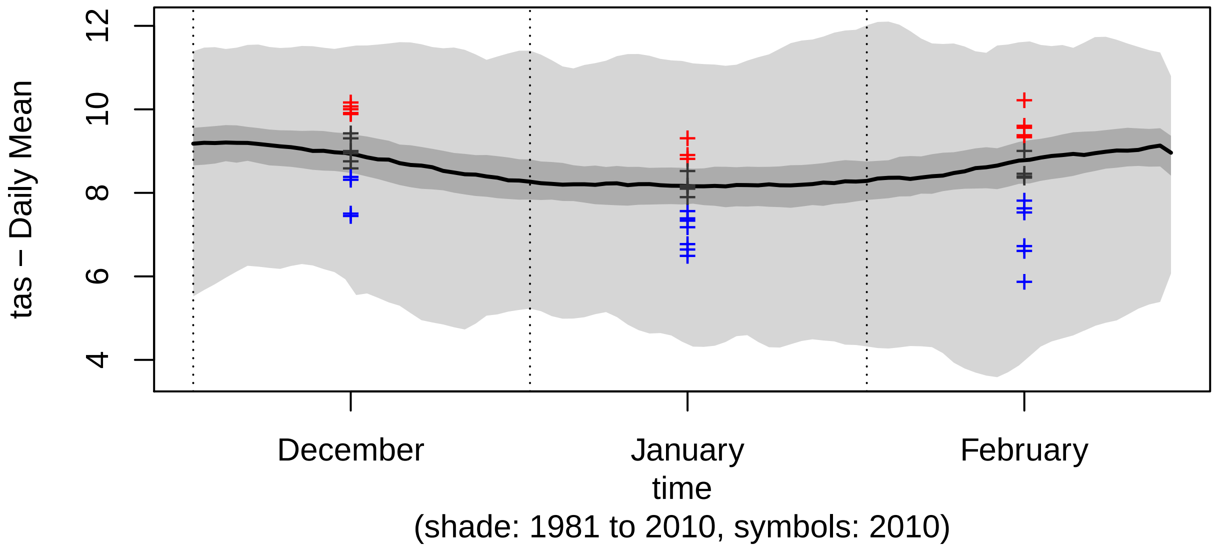



```r
spreadPlot(prd, obs, year.target, violin=T, add.points=T, pch=21)
```



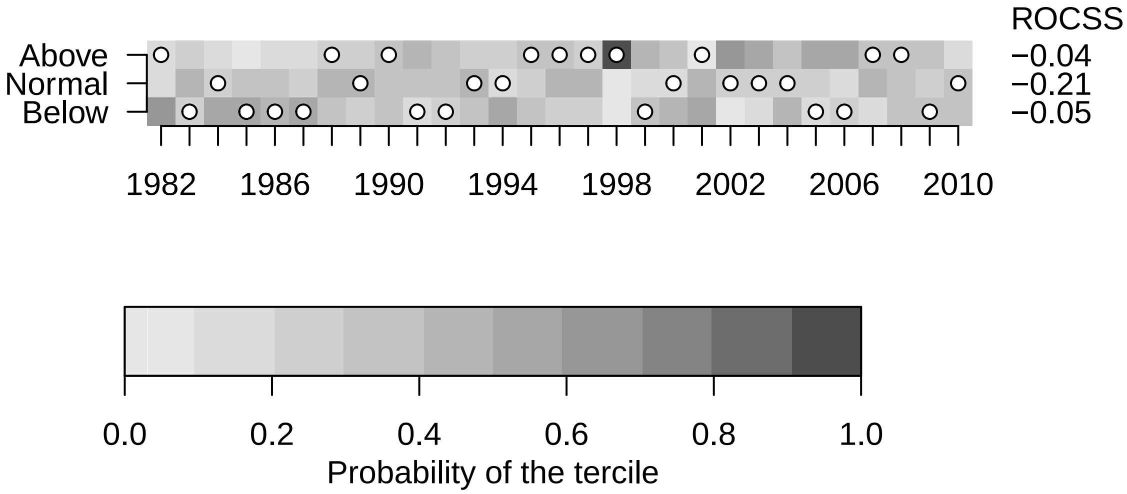
It is also possible to plot only the values of the ensemble members.

```
spreadPlot(prd, obs, year.target, boxplot=F, violin=F, add.points=T)
```

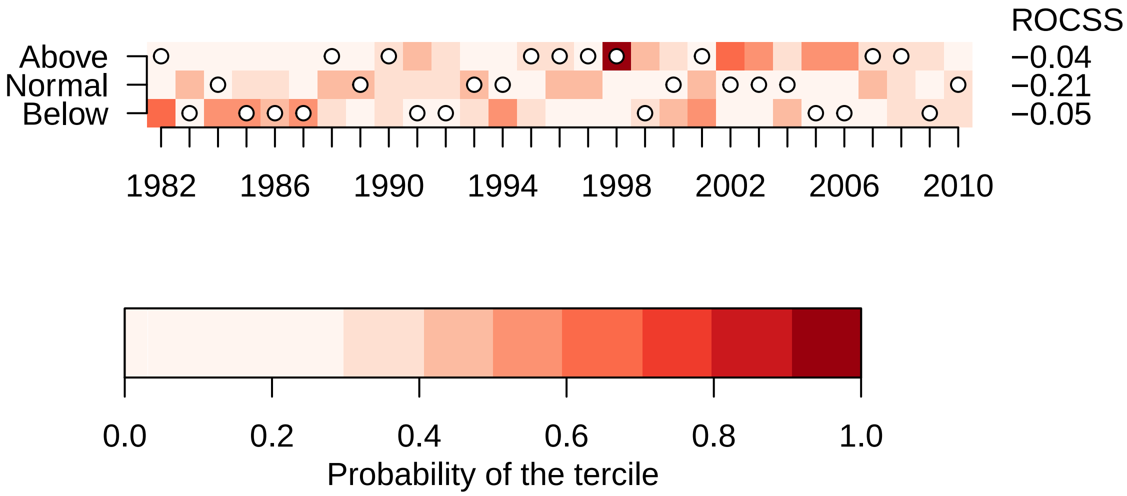


1.3.3 Tercile plot

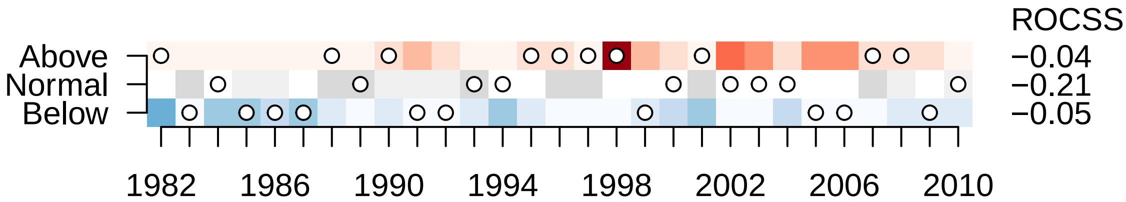
```
tercilePlot(prd, obs)
```

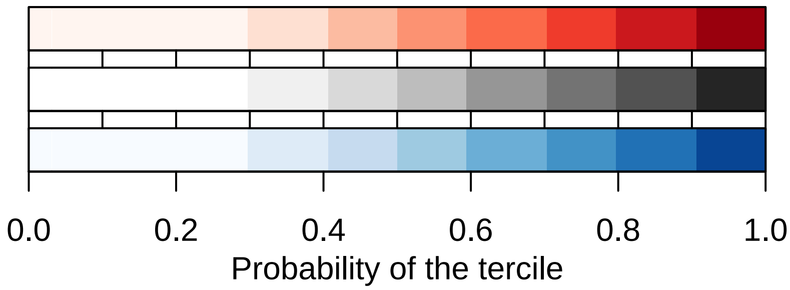



```
tercilePlot(prd, obs, color.pal="reds")
```




```
tercilePlot(prd, obs, color.pal="tcolor")
```



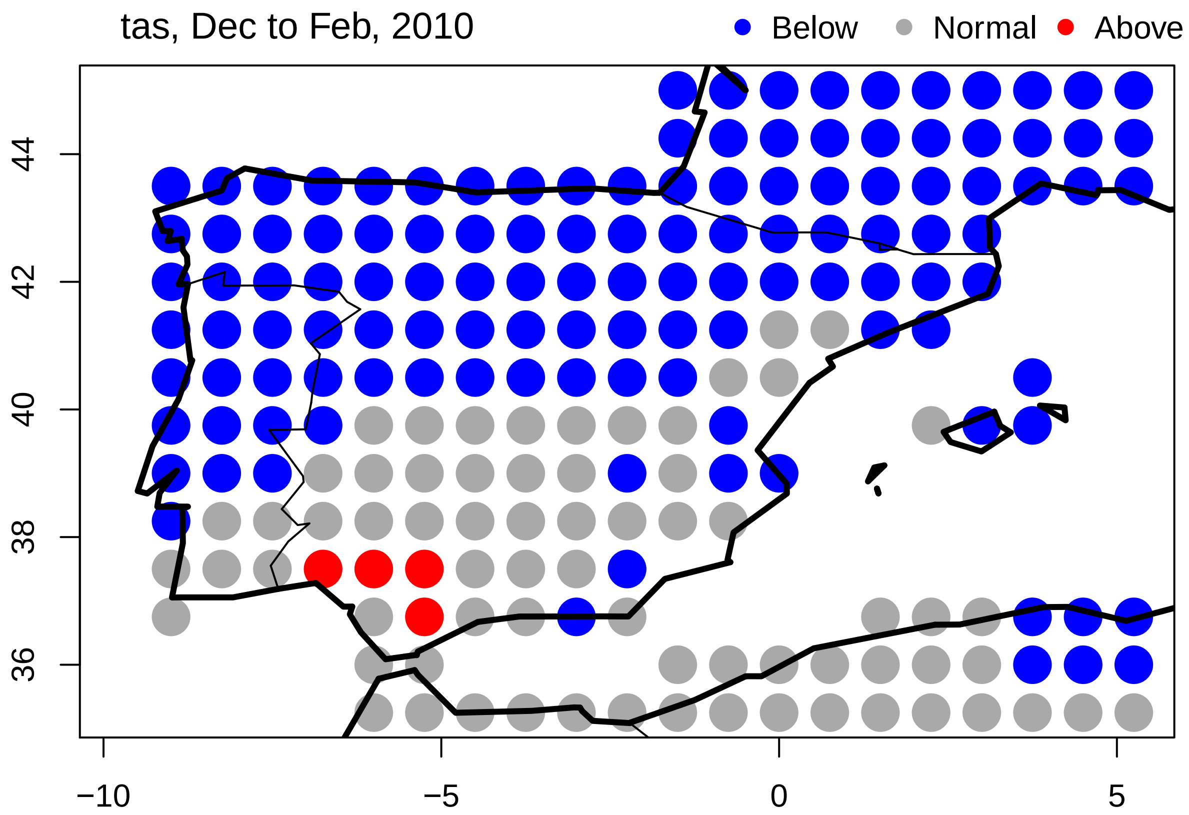


1.3.4 Bubble plot

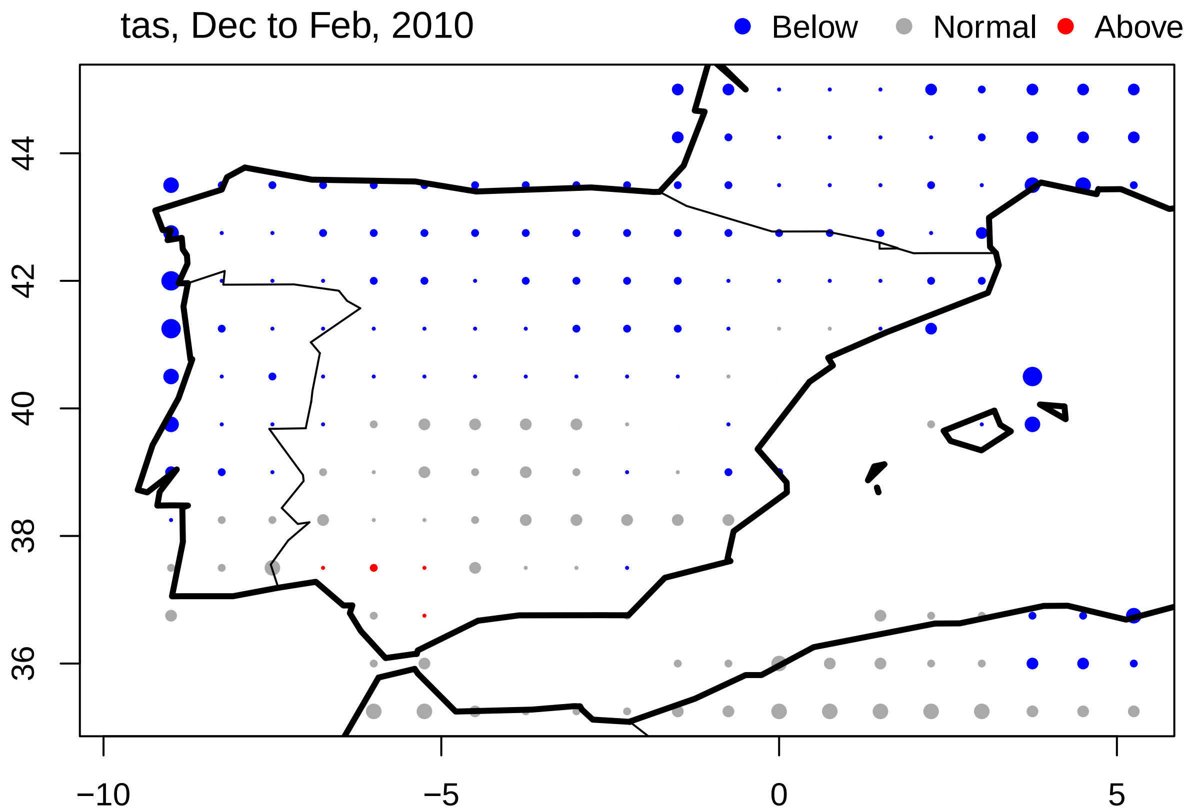
The bubble plot represents the most likely tercile in colors, the probability of that tercile with the size of the bubble (optional) and the skill of the forecast system for that tercile as transparency of the bubble (optional). Currently, the skill score used is the ROCSS. Only positive scores are shown (the negative ones –the system is worse than the climatology– are not plotted). Pie charts in stead of bubbles can be drawn indicating the predicted likelihood of each tercile.

The bubblePlot can be invoked with different levels of complexity:

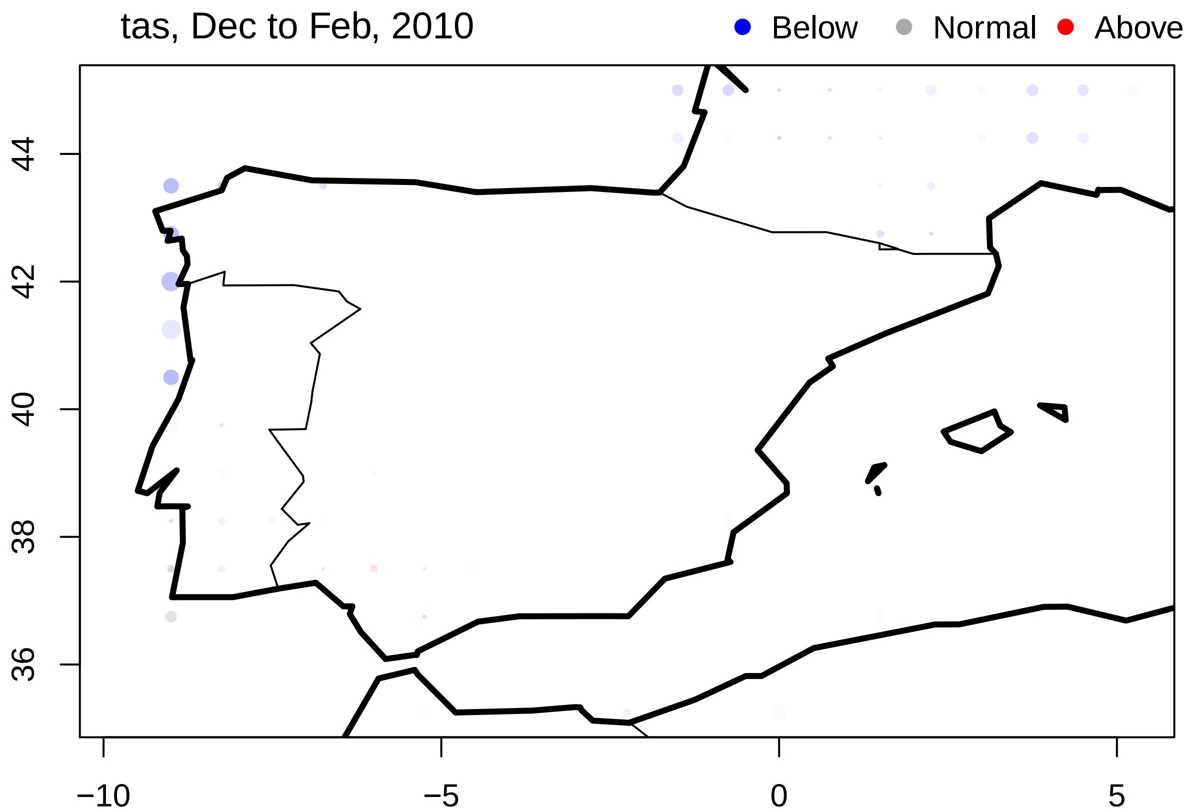
```
bubblePlot(prd, obs, year.target, size.as.probability=F, score=F)
```



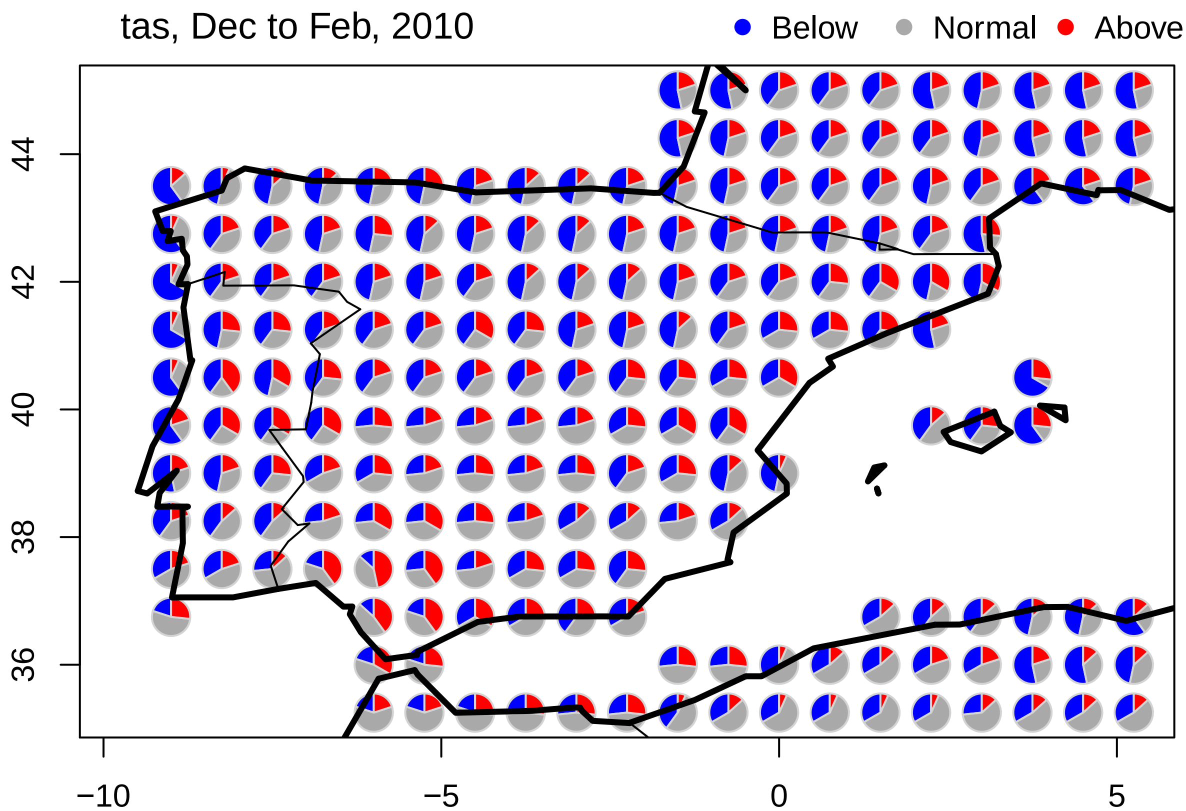

```
bubblePlot(prd, obs, year.target, size.as.probability=T, score=F)
```



```
bubblePlot(prd, obs, year.target, size.as.probability=T, score=T)
```




```
bubblePlot(prd, obs, year.target, pie=T, score=F)
```




```
bubblePlot(prd, obs, year.target, pie=T, score=T)
```