DTC visitor report
Sub-hourly variability of observations as a function of time-of-day, seasonal and geographical location

Marion Mittermaier

December 16, 2014

1 Background

In the UK 1-minute resolution synoptic observations are available across the observing network and this provides an opportunity for considering the within-hour variability and how this may be used in adding uncertainty to the hourly synoptic observations for their use in verification.

At present we treat synoptic observations as the absolute truth, though when one digs a little deeper into the way synoptic observations are taken (e.g. the WMO standards manual), and transmitted as SYNOPs, they are representative only in the broadest climatological sense. When these observations are used to verify NWP forecasts, they are therefore taken at face value and any error is wholly attributed to the forecast. Granted, many model variables are output as instantaneous time step values, which arguably forms yet another dimension to this error (and skill) attribution process. For this study 11 month time series for 7 locations around the UK are considered. These represent mixture of locations, from the south coast of England to Scotland, inland and upland sites.

In brief we want to:

- Identify a viable method for deriving observation uncertainty estimates from sub-hourly observations; and
- consider the time-of-day (TOD), monthly and seasonal dependence as a function of geographical location to determine how general (or specific) these uncertainty estimates are.

2 Data analysis

A stationary time series has a mean, variance, autocorrelation, etc. that stays constant over time. Meteorological time series are not that different from business or economic time series, as they are also often far from stationary when expressed in an untransformed state, e.g. temperature. Non-stationary time series typically exhibit trends, cycles, random-walking, and other non-stationary behaviour. Yet most statistical forecasting methods require stationarity and are based on the assumption that the time series can be rendered approximately stationary through the use of some mathematical transformation(s).

Temperature time series in particular exhibit strong trends and are strongly autocorrelated, the trend must be removed first, before the residuals can be analysed, the stationarity of the time series must also be tested for.

One of the methods tested is the fitting of a Generalised Additive Model (GAM) (Hastie and Tibshirani, 1986, 1990) as provided by the R package `mgcv`. It applied to the hourly synoptic data using a smoothing function. The GAM fit parameters were then applied to the 1-minute time series to create a time series of the same length for trend removal.

There are several benefits to transforming time series to achieve stationarity. One is being able to obtain meaningful sample statistics such as means, variances, and correlations with other variables. Such statistics are useful as descriptors of future behaviour only if the series is stationary.

The "method of moments" is used to fit the Pearson (Pearson, 1895) family of continuous probability distributions. The Pearson system was originally developed to model visibly skewed observations where the first two moments are not sufficient to capture the skewness or peakedness (kurtosis) of observed data sets. The family has seven classes.

I Beta, continuous uniform, normal in the limit

II Symmetric beta

III Chi-squared, exponential, gamma, normal in the limit

IV Cauchy, hypergeometric, normal in the limit

V inverse Chi-squared, inverse gamma, normal in the limit

VI Beta prime, F, normal in the limit

VII Student's t

Fitting a distribution to the observations gives a powerful parametric tool to support a limited observations data set to further understand observed behaviour. It can also be used to estimate particular distribution quantiles based on specific probabilities.

For the 7 sites and 11 months the overwhelming majority of monthly distribution of sub-hourly temperature ranges are Pearson I distributions (Beta), followed by type VI (Beta prime or F) and then by type IV (Cauchy). This is shown in Fig. 1. No other distribution types were identified.

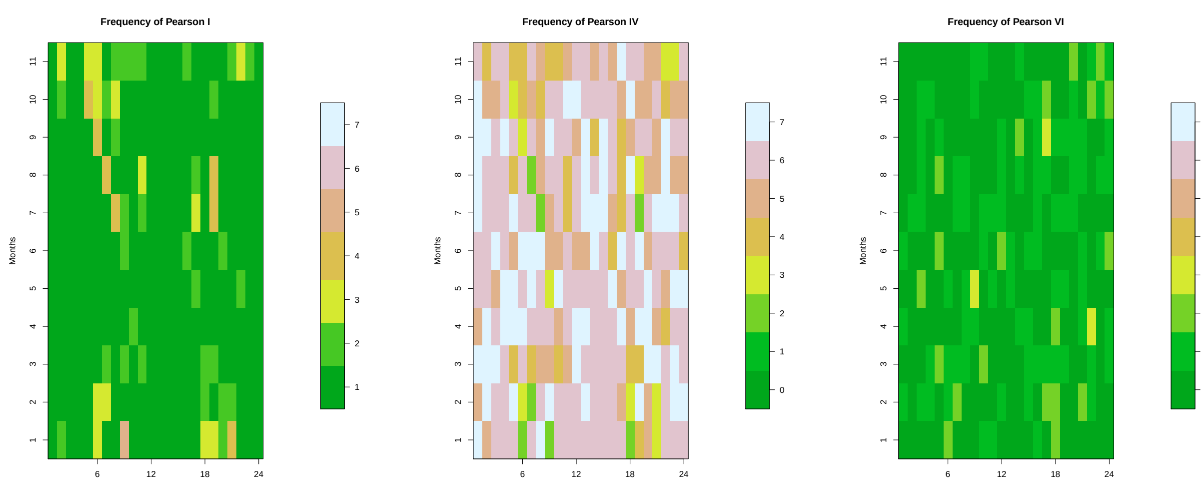

Figure 1: Prevalence of type I, IV and VI Pearson distributions as a function of time of day and month. Months run from August 2012 to June 2013.

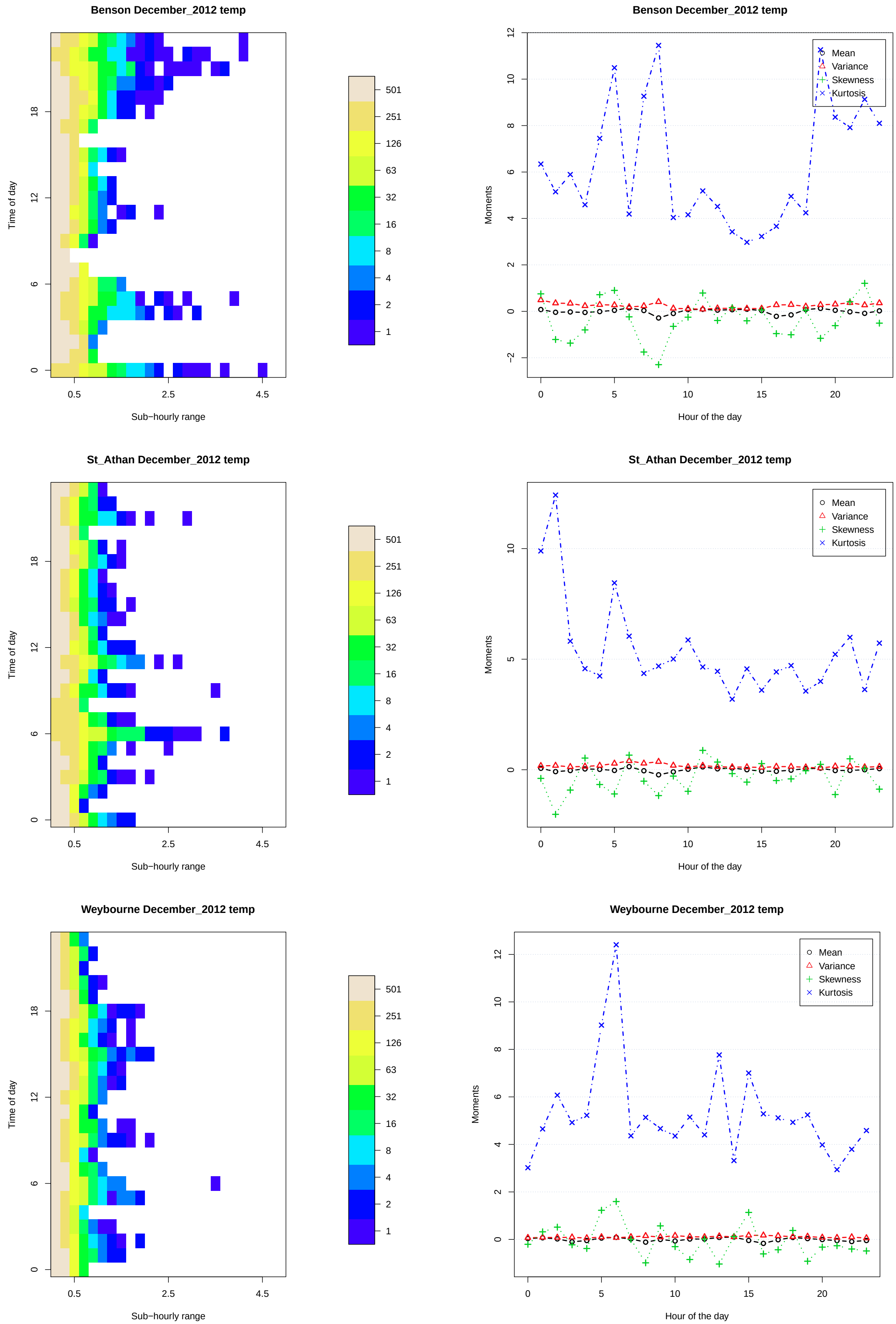

Figure 2: Sampled distributions (left) based on the calculated moments (left) as a function of time of day for three sites in December 2012.

Figure 2 shows three examples of the hour-by-hour moments and the resulting sampled distributions (1000 samples). The moments confirm that the trend removal has been largely successful. Benson is an inland site, whereas St Athan and Weybourne are coastal sites. Generally at least half the residuals are smaller than 0.5K, though Benson does exhibit a broadening of the residuals, particularly at night, which may be related to cold radiation night temperature fluctuations. The kurtosis of the residuals is highly variable as a function of TOD and locations.

Figure 3 shows three different sites for April 2013. There is a an increase in range during the middle part of the day. St Marys on the Scilly Isles off the SW tip of Cornwall shows the smallest residual range across the day, demonstrating the strong maritime influences this site experiences.

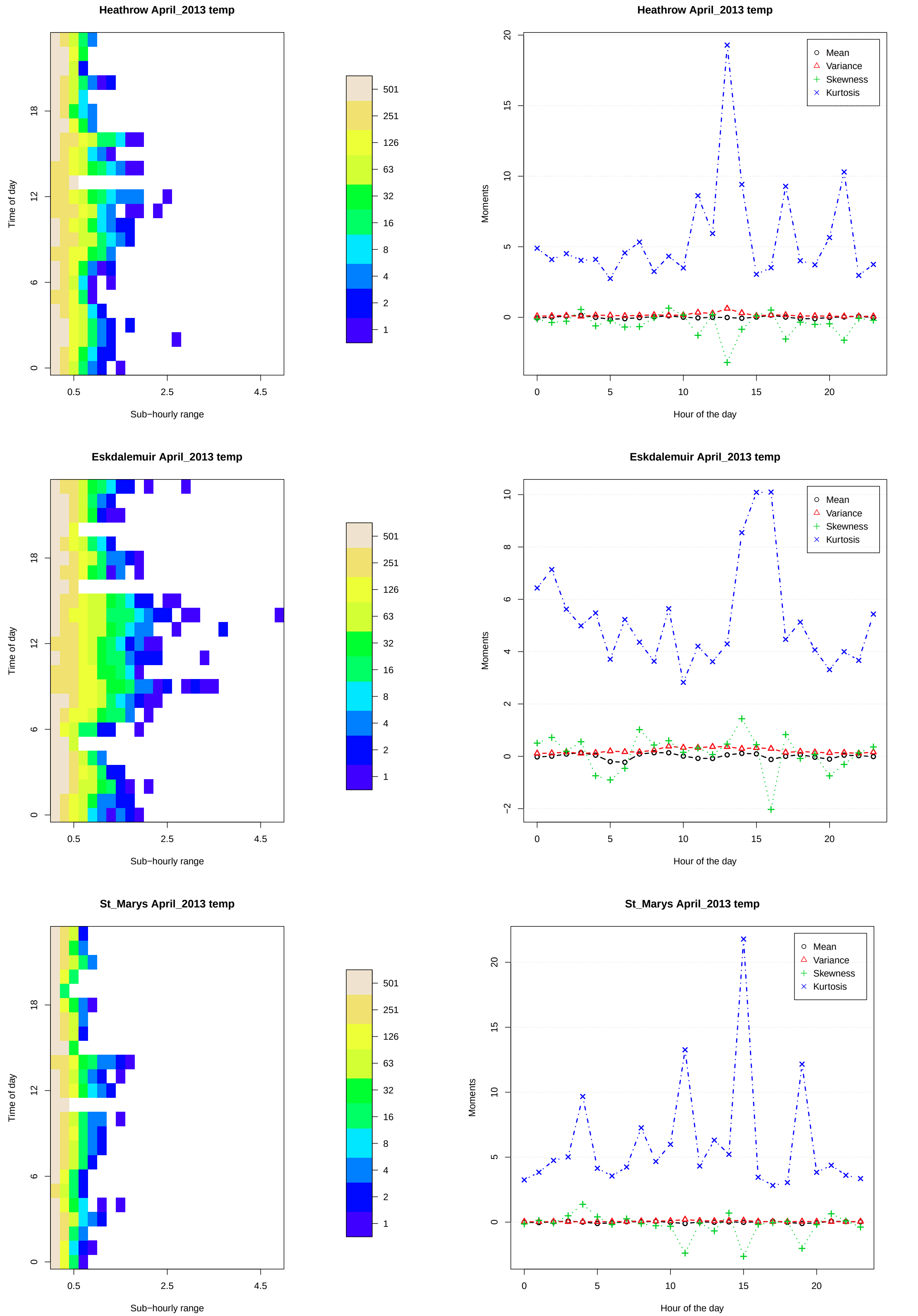

Figure 3: Sampled distributions (left) based on the calculated moments (left) as a function of time of day for three sites in April 2013.

The derived Pearson distributions were used to extract the quantile $Q = inf\{x \in R\colon p \leq F(x)\}$ with p=0.5. These are plotted in Fig. 4 for six of the sites as a function of time-of-day and month. There are similarities between all the sites, but also plenty of differences. South Uist (Scottish island) suggests the least systematic behaviour as a function of TOD or time-of-year. The other coastal sites show a more muted pattern with St Mary's being the most muted, and probably the most exposed to the maritime environment. Inland and upland sites are also similar, in terms of the distribution.

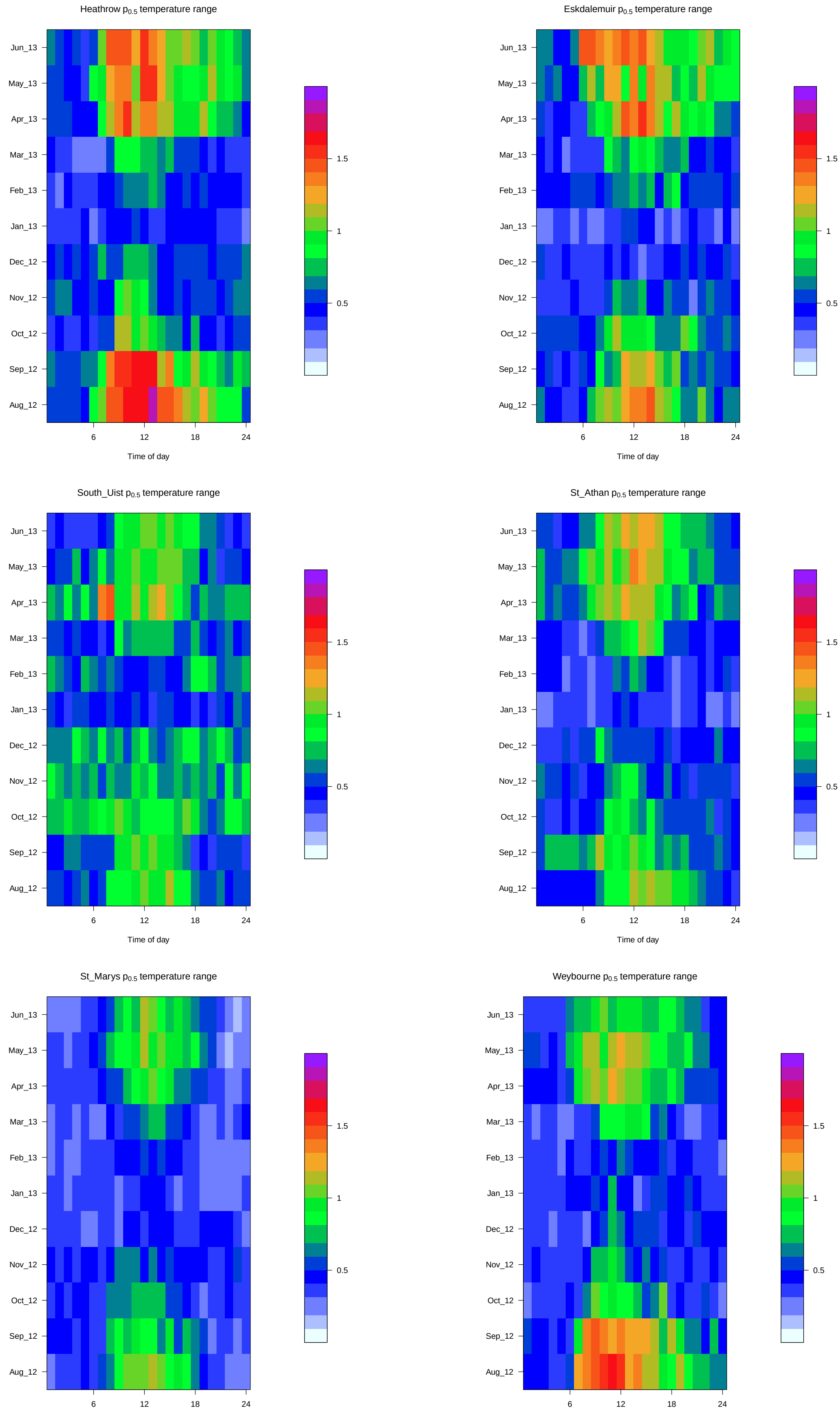

Figure 4: Derived quantile values for a 50% probability of occurrence as a function of time of day and month.

3 Concluding remarks

Median residual values from the sampled (simulated distributions) of within-hour temperature fluctuations suggest that values in excess of 0.5–1 K are not uncommon. Given that the quoted accuracy of many NWP model temperature forecasts (verified using hourly synoptic observations) is of this order, what is the impact (if any) of the sub-hourly fluctuations, and should these results affect the interpretation of verification scores calculated using hourly synoptic temperature values? This is the subject of ongoing investigation.

References

Hastie, T. and Tibshirani, R. (1986). Generalized Additive Models. *Stat. Sci*, **1**(3), 297–318.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman & Hall/CRC.

Pearson, K. (1895). Contributions to the mathematical theory of evolution, II: skew variation in homogeneous material. *Phil. Trans. Roy. Soc.*, **186**, 343–414.