NEXT QUESTIONS: splines vs arima? Can splines handle autocorrln? Arima errors in splines? Gls in splines?

* Aaron said they sometimes use cubic splines, but difficult seasonality - Step changes rather than smooth seasonality. Will it be weekly? Monthly? Daily?
  + Cubic splines: <https://www.nature.com/articles/s41409-019-0679-x> - The introduction has got a good rundown of these but essentially you divide the data into “knots” and then within each you fit a cubic polynomial with additional restrictions to ensure the polynomial points at the knot boundaries align. They have some good rules of thumb, EG first and last knot should be at around 5th and 95th percentile of observations; typically 5 knots is sufficient. Also mentions an R package for cubic splines.
  + Splines versus a loess smooth: second answer here is good. Suggests that for prediction splines are generally better, with “natural splines” doing well with far simpler calculations if the data set is very big. Also gives a link for “optimising the penalty value” for splines, whatever that means. <https://stats.stackexchange.com/questions/20032/comparing-smoothing-splines-vs-loess-for-smoothing>
  + **Excellent review of splines in general, and implementation in R.** Notes that what defines different types of splines is the order of the polynomial and the number and placement of knots. Gives examples of analysis on real data , fitting a spline within a GAM context, and also talks about penalised splines that have a tuning parameter that controls the degree of bendiness. All within R. Talks specifically about the value of penalised splines ; Basically you don't need to worry about the placement or number of knots as long as you start off with a very large number of knots and then let the penalty term prevent the overfitting. and there are functions that will easily do this within major R packages (see final paragraph before discussion). More info:
    - Splines can be divided into regression splines and smoothing splines . Smoothing splines are like loess, and are useful for complex patterns but not as easy to apply , whatever that means.
      * The main types of regression splines are B-splines and natural cubic splines. These are the two that the splines package uses.
    - <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-019-0666-3#Sec22>
  + Note on the mgcv package, in link above: “Mixed GAM Computation Vehicle with Automatic Smoothness Estimation, or mgcv is the package that offers many possibilities, has a large number of downloads and is currently supplied with the basic distribution of R. The package includes many different spline basis: thin plate regression splines, cubic regression splines and cubic regression splines with shrinkage, cyclic cubic regression splines and p-splines. All of these basis are relatively easy to use with a specification in the s fuction. The package performs automatic estimation of the smoothing terms and that makes it particularly useful in practice. It is important that the package is well documented and the help files provide enough details for the user to understand what is hidden behind the code. The package comes with several online material and a very well written book. There are just two sample datasets in mgcv, but the required nlme package also loads 41 datasets that can be used for better understanding of procedures. The added functionality of mgcv includes smoothing in two dimensions, allows the users to specify their own spline basis and also contains procedures for variable selection.”
  + Also: “the choice of model selection criteria should balance the competing objectives of conformity to the data and parsimony [1, 26]. This issue is closely related to the selection of a simpler or more complex regression model, for example by preferring AIC or BIC as the criterion for variable selection. Consequently, the default values of a spline package can be a very good choice to derive a model for one aim and a bad choice for another aim. These issues are more discussed in the context of variable selection, model complexity and model stability”
* This blog and the preceding ones are super useful! Forecasting energy demand based on hour and day effects, looking at seasonal decomposition with ARIMA (via running forecast() function on an stl() object), MR, GAM, GAMM with ARMA errors. <https://petolau.github.io/Analyzing-double-seasonal-time-series-with-GAM-in-R/> Also subsequent posts may look at boosting, RF, SVMs etc.
* For ARIMA vs seasonal decomposition: <https://stats.stackexchange.com/questions/396790/decomposition-of-time-series-vs-arima-model>. ARIMA and even SARIMA don’t have a decomp. A common analysis is ARMA on residuals of a seasonal decomp, e.g. STL (loess decomp). But also contains a reference to an R package that can do multiple-frequency decomp – useful for e.g. hourly data?
* The task will involve energy data (demand? Prices?) and forecasting with weather data. It’s nasty data that has all the problems they usually see
* Want to see lateral thinking about the problem
* In interview said small datasets are a problem, and often they have to use prior knowledge to pick how many predictors to put in a model
* Can have very low variation at a month level – MAE of 4 out of 100 – but can be MAE of 40 out of 100 at a daily level, because of human factors (‘trader having a bad day’ said aaron). Note this wasn’t specifically talking about ‘the task’ but was talking about one of their main models
* Good next steps for following modelling: If they ask , using prediction window as a tuning parameter. Is it best to predict using one month of data, one week of data, one year of data etc?
* Tidyverse forecasting: <https://robjhyndman.com/hyndsight/fable/>. Seems better than forecast package by a bit… see bottom of post for comparison. Also apparently it’s the successor to forecast, so future-proofed. Also here is the full tutorial, including an electricity demand one: <https://tidyverts.github.io/tidy-forecasting-principles/methods.html>. And available models (Currently not many) here <https://stats.stackexchange.com/questions/472610/r-forecasting-list-of-supported-or-available-models-from-fable-package>.
  + Vignettes:
  + <https://cran.r-project.org/web/packages/fable/vignettes/fable.html>
  + <https://cran.r-project.org/web/packages/fable/vignettes/transformations.html>