Time Series Forecasts with Random Forest

The questions:

* How to use random forest with time series?
  + Is it possible to overfit?
* How to tune the parameters?
  + Feature fraction
  + Minimum size vs. other criteria (for leafs)
* Applications
  + Significant variables (and how these change with horizon)

For tuning:

* Pick a few hyperparameters
* Select values for those hyperparameters
* Use cross-validation to choose best values

Random forest and cross-validation are inherently cross-sectional techniques. It doesn’t quite make sense to k-fold the time series data, because you’ll end up validating past data on future data. This seems problematic because the future data theoretically has no effect on the past data. For all we know, the time series is fundamentally different in the 2010s as opposed to the 1960s. Maybe there’s a structural break, maybe there’s some kind of non-linear time trend; at any rate, it would be best to avoid validating past predictions on future data.