Use of forecasting models for improved PID control in wastewater treatment

Kathryn B. Newhart

Table of Contents

##### Page break

# Introduction

One of the greatest and costliest threats to surface waterways in the United States is nutrient pollution (Environmental Protection Agency) 2011). The removal of nitrogen and phosphorus from water at municipal wastewater treatment plants (WWTP) is necessary to address the growing environmental challenge. As states implement increasingly stringint nutrient limits on WWTP, the cost of has wastewater treatment increased substancially. Biological nutrient removal is acheived by introducing dissolved oxygen (DO) via industrial air blowers, which is one of the largest operating costs for WWTP. Conventionally, air blower output is increased or decreased relative to the measured concentration of DO in an aeration basin to maintain a DO setpoint where nitrogen removal is effectively guarenteed. While this approach ensures treatment, there are multiple drawback. DO control is energy inefficient as aeration is provided whether ammonia is present or absent; conventional control leads to large fluctuations in air demand to maintain a single DO setpoint, which mechanically stresses the air blowers; and microbial communities responsible for treatment are stressed with large and rapid DO fluctuations. Currently, WWTP operators rely on trial-and-error to identify a DO setpoint that treats water to an ammonia concentration well below the permited limit. Such a large factor of safety is required to ensure that under the highest flow and load conditions, water will be fully treated and the WWTP will not exceed their nutrient limits. To save energy and reduce excessive use of air blowers at WWTP, a new aeration paradigm is needed.

Ammonia-based aeration control (ABAC) is an aeration approach which responds to water quality changes in real-time by adjusting air flow to meet an ammonia setpoint instead of a DO setpoint. ABAC can limit excessive aeration promoting full nitrogen removal, potentially improving phosphorus removal, and reducing effluent ammonia peaks (Rieger et al. 2012). The most popular ABAC configuration utilizes a microbial kinetics model to calculate the aeration demand given a variety of operating parameters such as solids retention time and influent ammonia concentration (Duffy et al. 2010). This feedforward approach can reduce aeration demand, but requires extensive sampling and specialized knowledge for accurate model calibration and implementation. A feedback ABAC option would not require a kinetic model and could adjust DO setpoints from an ammonia sensor in the aeration basin where treatment takes place. However, WWTP operators are hesitent to adopt ABAC due to the instability of the sensors involved in feedback ABAC (specifically ion-selective electrodes) and the potential for the growth of microorganisms that inhibit settling in subsequent treatment at lower DO concentrations (i.e., filamentous bacteria). A third ABAC approach could incorporate statistical process models to predict ammonia in a feedback or feedforward configuration. The advantages of forecasting using statistical and machine learning models is (a) no additional sampling, microbiological analysis, or proprietary software is required to build the model and (b) the forecast can easily replace the current measured value of ammonia in the supervisory control and data acquisition (SCADA) system of the WWTP–which lacks advanced control schema. Existing hurdles to implementation of data-driven systems in WWTP include knowledge and experience. To date, there are only a handful of WWTP in and outside of the U.S. who have implemented model predictive control systems in their facilities.

The goal of this work is (1) to demonstrate the stability of a feedback ABAC configuration for WWTP operators and (2) to explore simple, data-driven methods of forecasting ammonia to overcome the lag associated with conventional feedback control to improve accuracy and reduce mechanical wear of aeration systems. Four different cascade control aeration configurations are tested and compared from a medium-sized WWTP to identify the most robust aeration control to balance treatment of ammonia with energy consumption. The manuscript is organized as follows: (1) an introduction to control systems in WWTP, (2) summary of methods for quantifying variation in multivariate systems, (3) summary of staistical and machine learning methods used to build the ammonia forecasting models, and (4) an assessment of how forecasting models can improve conventional control in WWTP.

## Control in wastewater treatment

WWTP are similar to other industrial processes in that select monitored system parameters need to be within a set range for the system to operate properly. Unlike many industrial processes, municipal WWTP have little control over the quantity and quality of the inputs to their system but are required by law to maintain a certain quality of the output. Due to the wide variation, manual adjustments of an open-loop control system (i.e., constant control output regardless of system conditions) cannot constantly achieve the level of treatment needed; frequently under-treating during peak flows and potentially exceeding regulated quality limits, and over-treating during low flows which wastes energy and other material inputs. Therefore, a flexible and responsive control system is required to maintain effluent water quality while minimizing energy and chemical input.

The most common control structure is feedback control in industrial automation. Feedback control determines a control action from a process measurement within the system (i.e., closed-loop control), and is able to automatically respond to system disturbances without specific knowledge of how the control and response parameters are related. The Proprotional-Integral-Derivative (PID) controller is the most common feedback controller due to its simplicity and robustness to respond to a deviation from desired conditions (i.e., error ). A PID control action is the sum of: a proportional term () where is a constant value; a integral term () and incorporates past control error with the integral function; and a derivative term () which anticipates future error with the derivative function. In the wastewater treatment industry, the derivative term is frequently set to 0 (i.e., PI control) due to the amplification of noise in the measured variable (Visioli 2006).

A single PID control loop can address straightforward problems, such as adjusting the speed of a blower or pump relative to a sensor measurement. To address more complex, nonlinear problems, multiple PID control loops can be combined in series to form a cascade control structure(Brosilow and Joseph 2002). In the case of ABAC, an ammonia sensor and setpoint define the outer/master control loop while inner/slave control loops define control variables such as DO, air blower flow, air blower speed, etc. In controller design, actuator and sensor dynamics and wear-and-tear are frequently ignored (Visioli 2006). For aeration at WWTP, this leads to a delay between a change in demand and the aeration provided, and excessive ramping of the air blower as a result of the delay.

What other topics need to be covered in the introduction?

# Methods and Materials

## Boulder Water Resource Recovery Facility

The Boulder Water Resource Recovery Facility (BWRRF, Boulder, Colorado, USA) is a 25 million gallon per day (MGD) municipal WWTP, currently operating a Four-Stage Bardenpho process at an average of 12 MGD. Given the high altitude of the facility and low daily ammonia limits (>1.9 mg/L NH4 as N), oxygen transfer efficiency is relatively low and results in high aeration demand (Burger et al. 2019). The current DO control configuration frequently over-aerates causing conditions that inhibit full nitrogen removal (i.e., denitrification in downstream anoxic zones). Consequently, aeration accounts for 35 to 50% of BWRRF’s energy consumption. There are multiple aeration control methods programmed into BWRRF’s supervisory control and data aquisition (SCADA) system, all of which rely on cascade control: airflow, DO, and ABAC. Airflow produces a constant volume of air by adjusting valves at the inlet of the aeration basin, regardless of the air demand; DO adjusts the volume of air to acheive a DO setpoint; and ABAC adjusts DO setpoints to acheive an ammonia setpoint (Figure 1). The process variables included in the control logic (), stability analysis, and forecasting model are located in one of BWRRF’s three aeration basins (Figure 2).

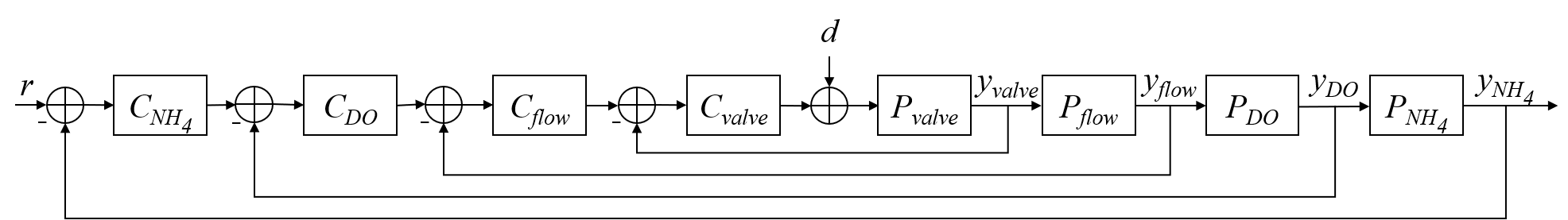


Figure 1. Cascade control logic for ABAC at BWRRF. The reference signal is the ammonia setpoint. The difference between and the actual ammonia measurement is input to the ammonia PID controller which adjusts the DO setpoint. The difference between the DO setpoint and the DO measurement is input to the DO PID controller which adjusts the air flow setpoint. The same logic flows for the air flow controller and air valve controller , which ultimately provides a certain volume of air to the aeration basins. A disturbance , such as a change in water quality, will impact the processes within the basin as measured by . The difference between the setpoints and values will force the PID controllers to continuously provide adjustments within their loop until all setpoints are met.

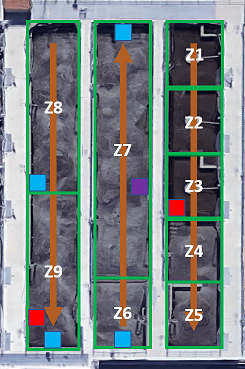


Figure 2. Flow, zone, and sensor diagram of one of the activated sludge aeration basins at the BWRRF. Orange arrows indicate the direction of flow through the basin. Zones 4-8 are aerated and have individual DO sepoints for each zone. Red boxes indicate nitrate sensors, blue boxes indicate DO sensors used in DO control, and the purple box indicates the ammonia sensor used in ABAC control.

The DO concentrations in Zones 6, 7, 8, and 9 of the aeration basins were continuously monitored using Endress Hauser (Reinach, Switzerland) COS61D optical DO sensors. In these zones, blower flowrate and valve position are also monitored and recorded in SCADA. AmmoLyt® Plus 700 ion-selective ammonia sensors from YSI (Yellow Springs, OH) are located in the aeration basin influent channel and in all three aeration basins in Zone 7. Ion-selective nitrate/nitrite sensors from YSI are located in Zones 3 and 9 of each basin; and the effluent channel after the final polishing aeration basin has an AmTax and NitraTax online analyzer (HACH, Loveland, CO). Aeration basin influent flow rates, wastewater temperature, and pH of the plant influent were also monitored. Online sensors were regularly maintained and calibrated by operations staff and readings periodically compared to laboratory results.

Data are collected and managed in the GE Proficy® system. For analysis, data was exported in 5-minute intervals into Microsoft Excel and imported to the statistical platform, , for analysis. Observations that were identified as “Bad” within the Proficy system (i.e., due to sensor calibration or during power loss) were not used. For QA/QC, all Proficy exports were kept in their native format.

The control configurations analyzed are outlined below (Table 1). A conventional DO aeration control method with a DO setpoint of 2.5 mg/L is considered the experimental control. Three ABAC configurations were tested, including 3.5 mg/L and 4.0 mg/L setpoints (at the middle of the aeration basin) and a 90 second and 300 second time delay between control updates.

Table 1. Control configurations tested vary by which is the “control” variable (i.e., outer/master loop in cascade control), the master setpoint for the control variable, and the delay term for the master PID loop

|  |  |  |  |
| --- | --- | --- | --- |
|  | Control Variable | Setpoint (mg/L) | PID delay |
| Control | DO | 2.5 | 90 |
| Test 1 | Ammonia | 3.5 | 90 |
| Test 2 | Ammonia | 4.0 | 90 |
| Test 3 | Ammonia | 4.0 | 300 |

## Stability assessment

The goal of the stability assessment is to provide WWTP operators with a quantative metric to decide between two control configurations on the basis of variability. To measure the stability of each operating configuration holistically, we investigated two multivariate metrics of variance: Total Sample Variance (TSV) and Generalized Sample Variance (GSV). TSV is the trace of the variance-covariance matrix of variables () and does not account for correlation between process variances. GSV is the determinant of the variance-covariance matrix () and will be small when there is a strong correlation between process variables.

To compare TSV and GSV values from a pair of control configurations, the difference between and the ratio of the values were calculated. The hypothesis tests used to determine if the difference in TSV or GSV are statistically significant are:

To determine if the difference or ratio was statistically significant, a Monte Carlo simulation was used. Data from the control and test conditions were combined into a single dataset, and then randomly split into two datasets 1000 times–this represents the null hypothesis in which there is no difference between the two randomly selected datasets. The actual metric is compared to the 1000 simulated metrics by the percentage of simulations that are greater than the observed (i.e., value). If the difference is considered statistically significant (), then the test control scheme is said to be less variable and more stable than conventional DO control.

## Ammonia forecast

To train and test the forecasting model, the data must be aligned to simulate real-time prediction. For a dataset with *n* rows, observations ) of all monitored process variables () will be merged with ( observations of the forecasted variable to create a matrix with columns and rows. Models are built using 1-6 days of process data and tested on the remaining days. Given the shortest dataset provided by BWRRF was 7.4 days, all datasets were shortened to only include 7.4 days of continuous, 5-minute level observations. The data compiled for the following analysis can be found in the [Appendix](#timeseries). All data was scaled to zero mean and unit variance for each control configuration.

To quantify model accuracy, the coefficient of determination (R2) and root mean squared error (RMSE) are used. R2 is a measure of the linear correlation between the actual forecast and model forecast values for the training dataset. An is a perfect model fit, while an indicates no correlation. RMSE is a weighted measure of average error and is used to compare model testing performance, with the goal of minimizing the RMSE. The percent improvement expected by use of the forecasting model in lieu of the current ammonia sensor in the cascade control configuration currently used by BWWRF is calculated from .

### Diurnal model

The purpose of incorporating a diurnal component into a forecasting model is to capture the time-dependent component of the response variable, in this case ammonia. While the daily trend of ammonia loading to a wastewater facility is acknowledged, it is rarely modeled Figure 3. The predictors in a diurnal model are various degrees (*n*) of sine and cosine functions where *t* is the minute of the day from 0 - 2, and are fitted linear model parameters:

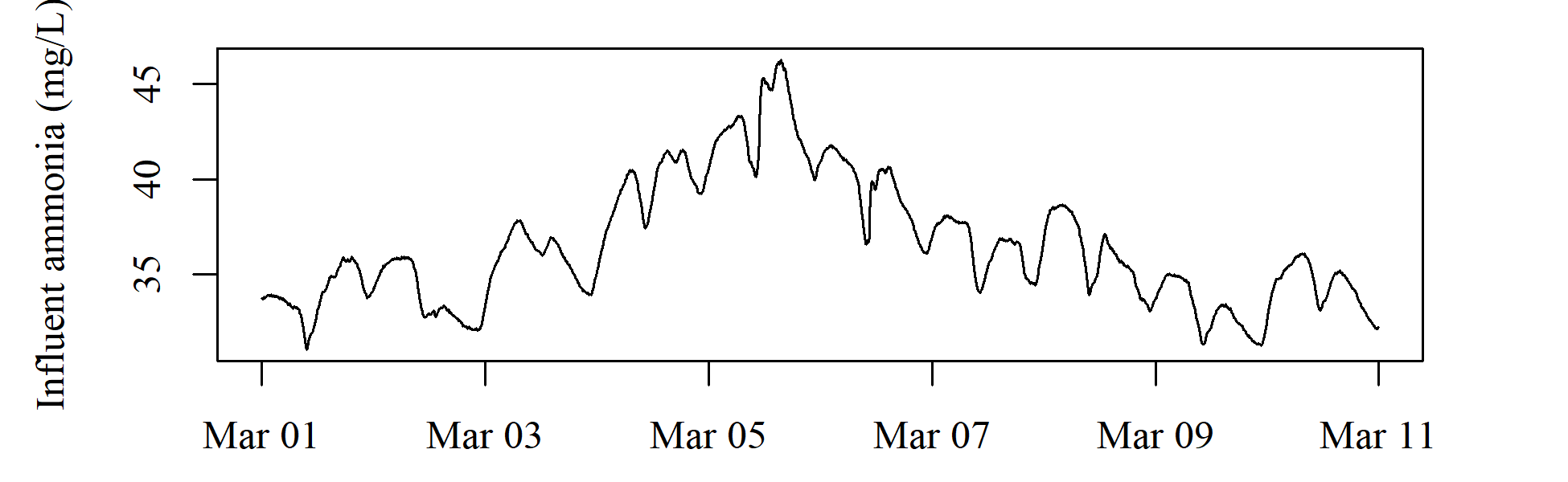


Figure 3. Timeseries plot of influent ammonia at BWRRF. Should we use the boxplot image or a timeseries to demonstrate a diurnal trend in ammonia?

#### Linear model

In a standard linear model, a response () is described using a set of predictor variables () and their corresponding model parameters where is an error term:

Typical linear models are fit using oridinary least squares (OLS), in which model parameters are estimated in such a way that the loss function that minimizes the sum of squares of the residuals, but prediction variability and model interpretabililty can be improved using alterntive fitting procedures (James et al. 2013). Lasso (or Least Absolute Shrinkage and Selection Operator) is one such regularization method that can decrease model complexity while maintaining or improving accuracy. In lasso, the sum of squares of the residuals are minimized using a loss function that additionally penalizes the size of model parameter estimates, shrinking insignificant predictors to zero.

The ‘shrinkage’ term responsible for driving the coefficients to zero is controled by . When , ridge regression returns the same linear model coefficients as OLS. The larger the value, the fewer predictors will be included in the model which decreases variance but increases bias in the response. Both the diurnal model and linear model were trained and tested using the glmnet package to find the best cross-validated .

### Machine learning

Neural network model?

# Results

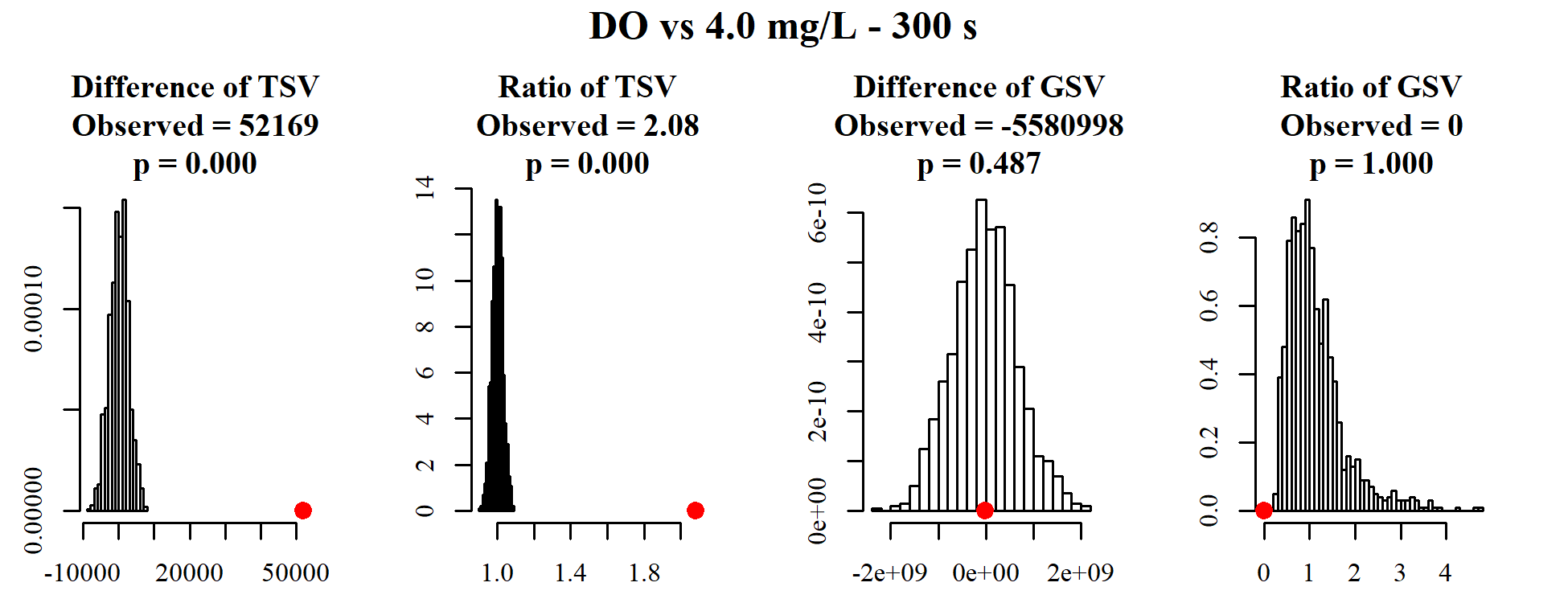
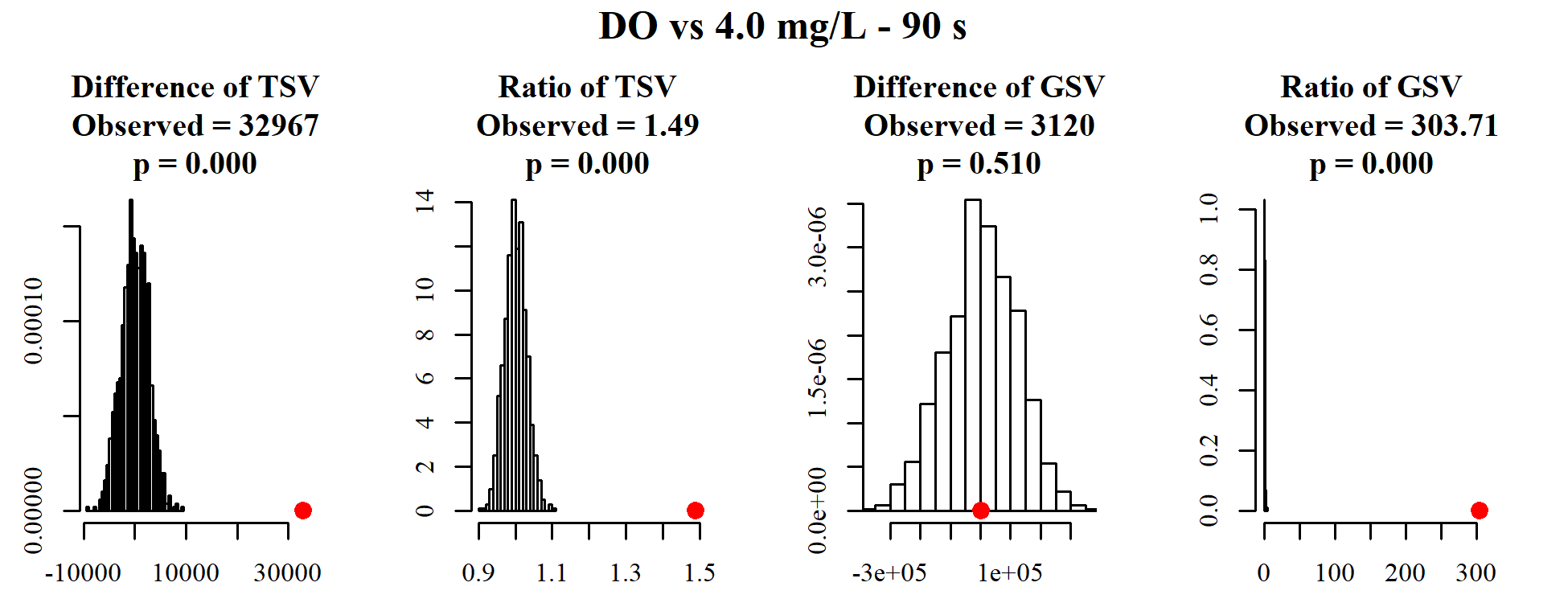
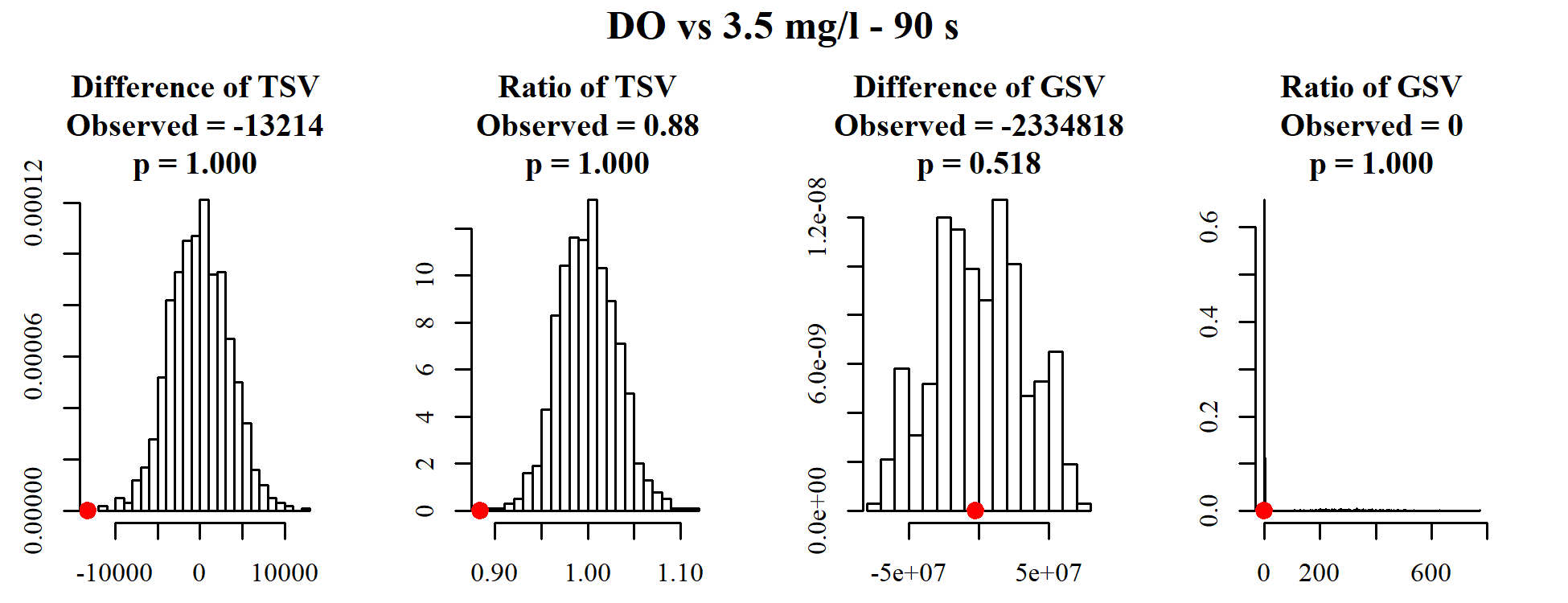
## Stability assessment

The results of the TSV and GSV calculations for each control configuration tested are in Table 2. The TSV and GSV for the control (DO control) condition are compared to other test conditions by subtracting the TSV and GSV, respectively, from the DO condition and by dividing the TSV and GSV, respectively, from the DO condition. To determine if the TSV and GSV diffrence or ratio are statistically significant, a Monte Carlo simulation is used (Figure 4). Both the difference and ratio of TSV values from two confiurations are considered stable metrics, as the historgram is approximately normally distributed. Whereas the difference and ratio of GSV values are frequently skewed and when the distribution is approximately normal, show no difference between test configurations. This demonstrates that GSV is an unable metric for comparing operating conditions. The large difference between the Monte Carlo simulation of the combined population and the observed TSV metric in the case of ABAC 4.0 mg/L 90 s and 300s verify that ABAC with a 4.0 mg/L setpoint is a more stable operating condition than DO.

Table 2. Control configurations tested vary by which is the “control” variable (i.e., outer/master loop in cascade control), the master setpoint for the control variable, and the delay term for the master PID loop

|  |  |  |
| --- | --- | --- |
|  | TSV | GSV |
| DO | 1.0e+05 | 3.1e+03 |
| 3.5 mg/l - 90 s | 1.1e+05 | 2.3e+06 |
| 4.0 mg/L - 90 s | 6.7e+04 | 1.0e+01 |
| 4.0 mg/L - 300 s | 4.8e+04 | 5.6e+06 |

##### Page break

 Figure 4. Histograms of the difference or ratio of test configuration’s TSV and GSV values compared to the control configuration (DO control). For example, or . Red dots indicate the observed TSV or GSV difference or ratio for the given test case. *p* values are calculated by the percentage of simulated case metrics that are above the observed metric.

## Ammonia forecast

### Diurnal model

The initial diurnal model fit used a single sine/cosine pair. However, this approach did not capture all visible cyclic patterns. Further testing evaluated the model fit of 1 - 200 sine/cosine pairs. The best diurnal model fit for the training and testing data of each control congiuration was effectively achieved using a 6 degree diurnal model (Figure 5). The realively low R2 value of the ABAC 3.5 control configuration is evident of abnormal variation in the minimum and maximum ammonia values (Figure 6).

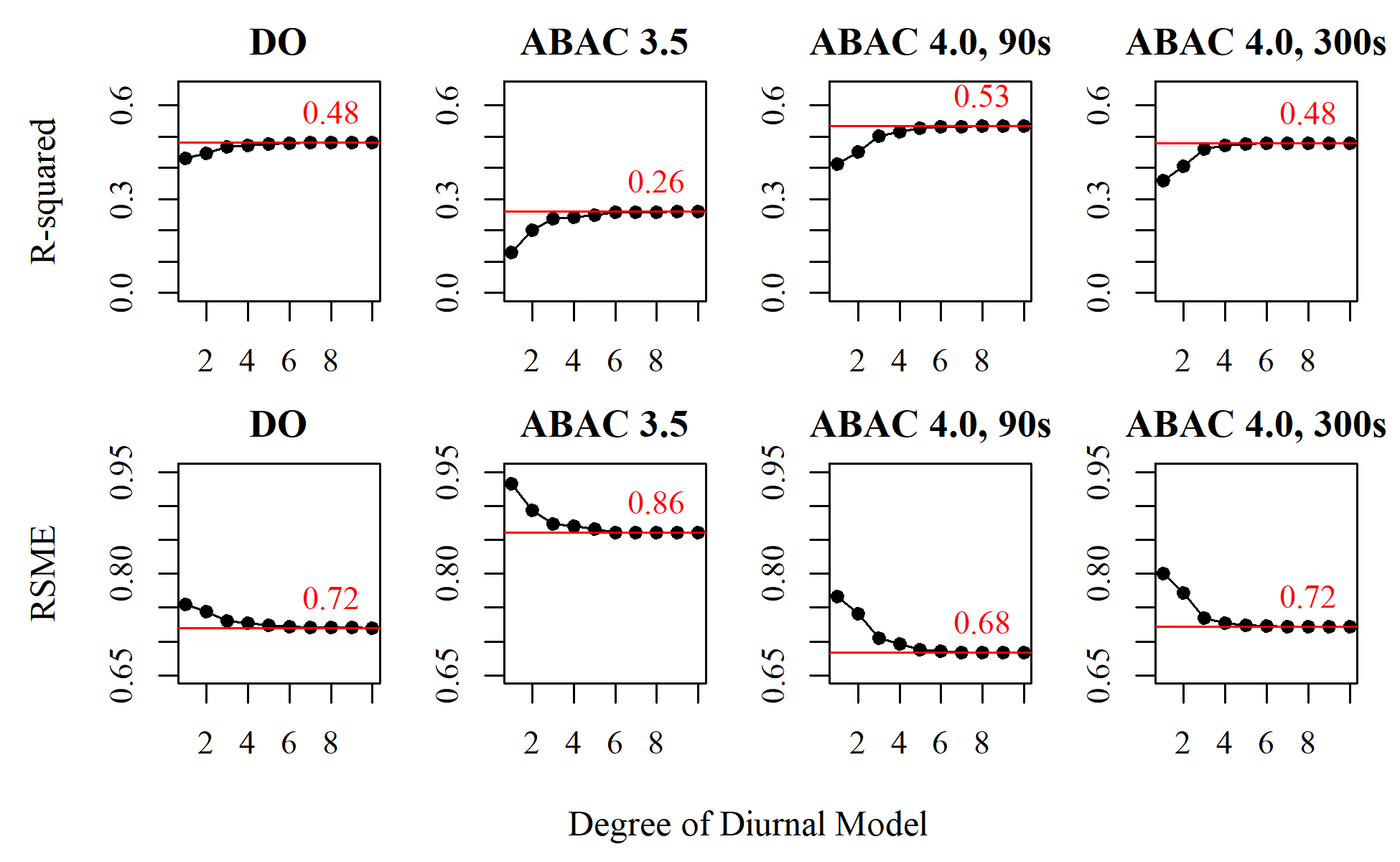


Figure 5. Diurnal model fit as function of degree for each control configuration. The red line indicates the R2 (top) or RMSE (bottom) value for a 10th degree diurnal model, which is effectively achieved by a 6 degree or fewer diurnal model.

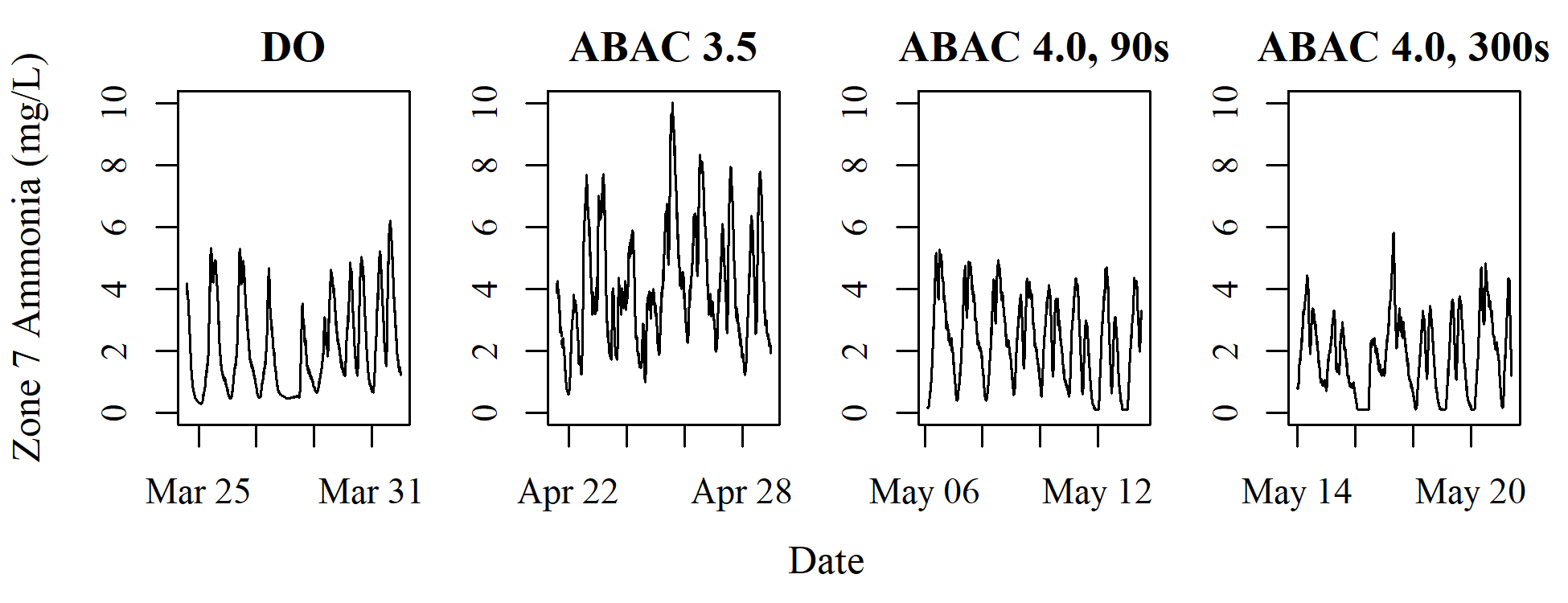
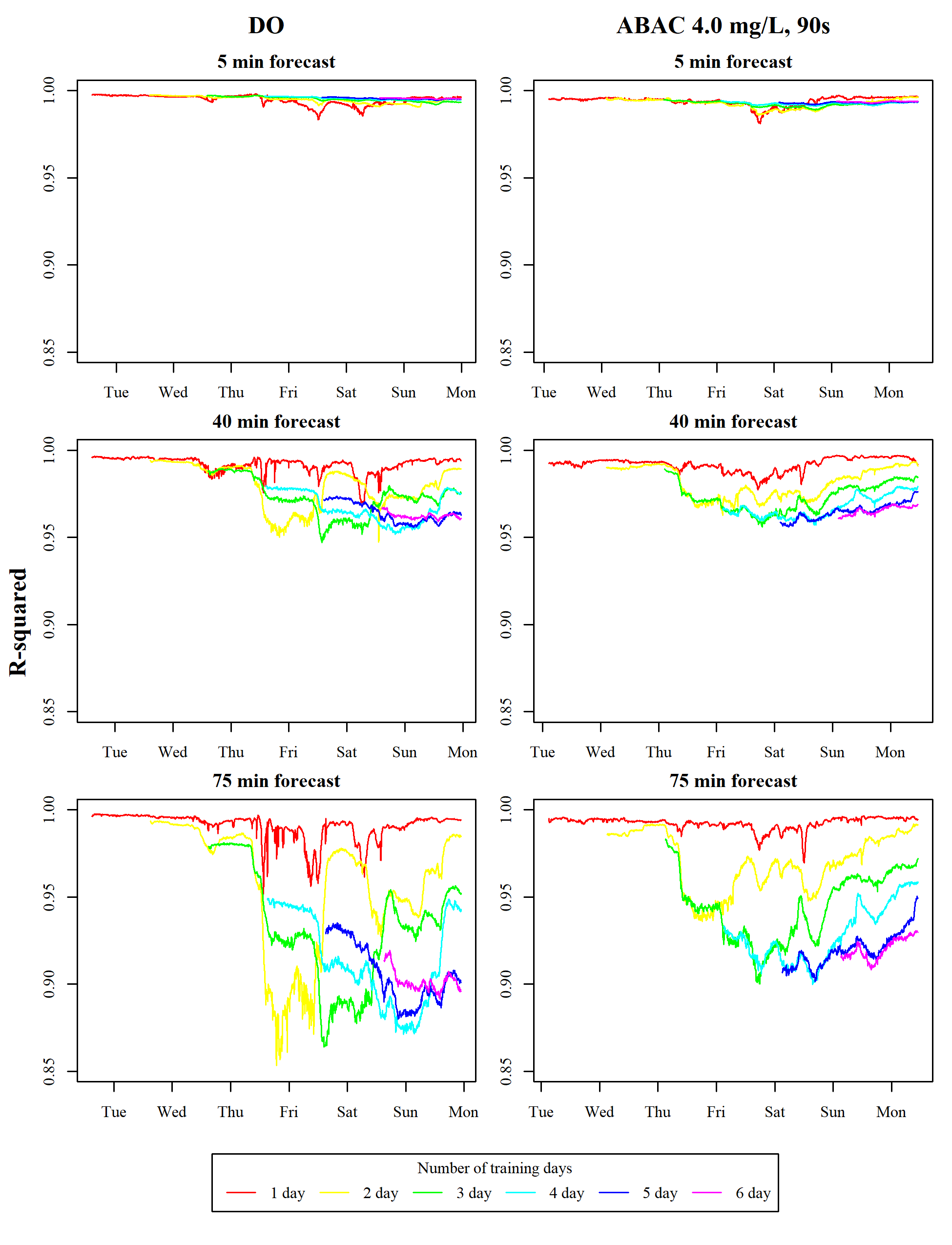


Figure 6. Timeseries plot of zone 7 ammonia at the Boulder Water Resource Recovery Facility.

### Linear model

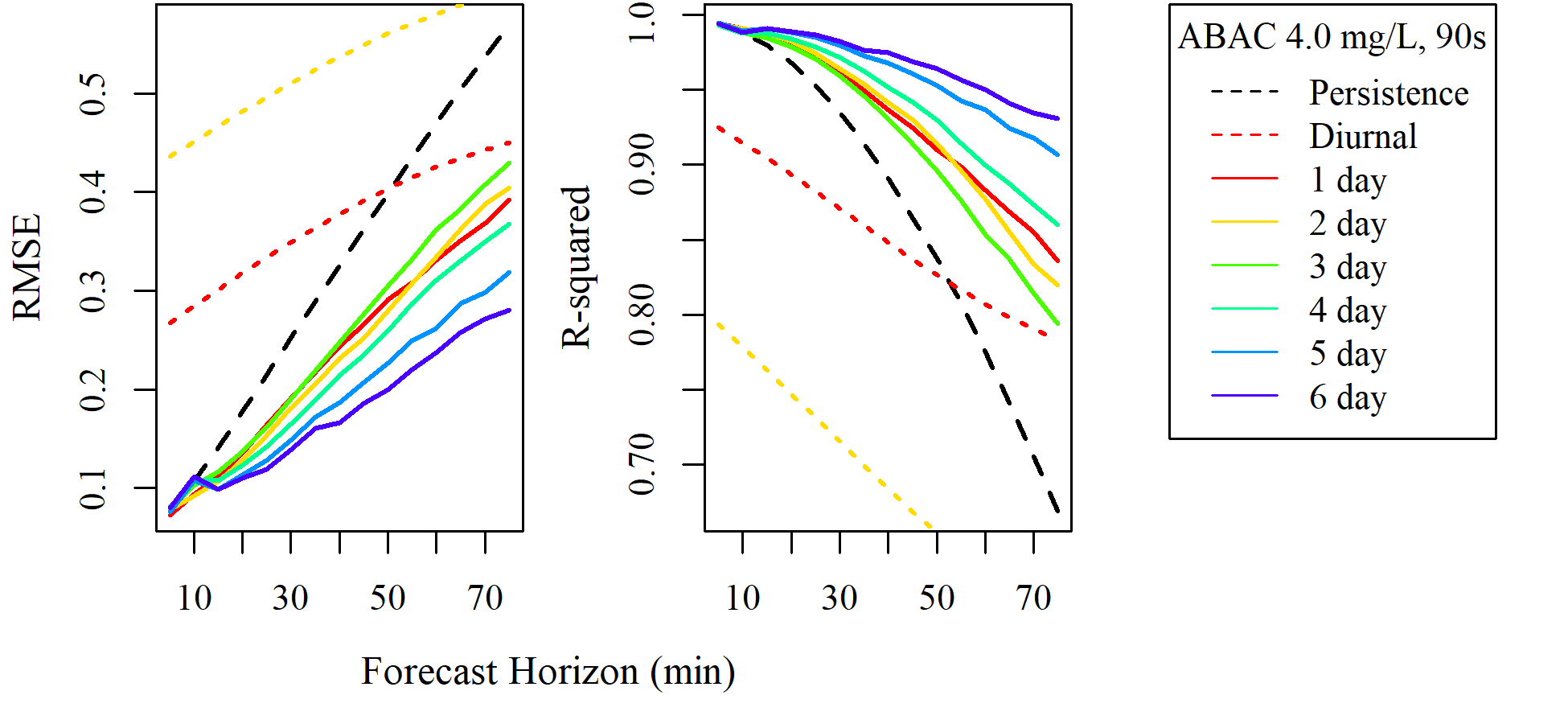
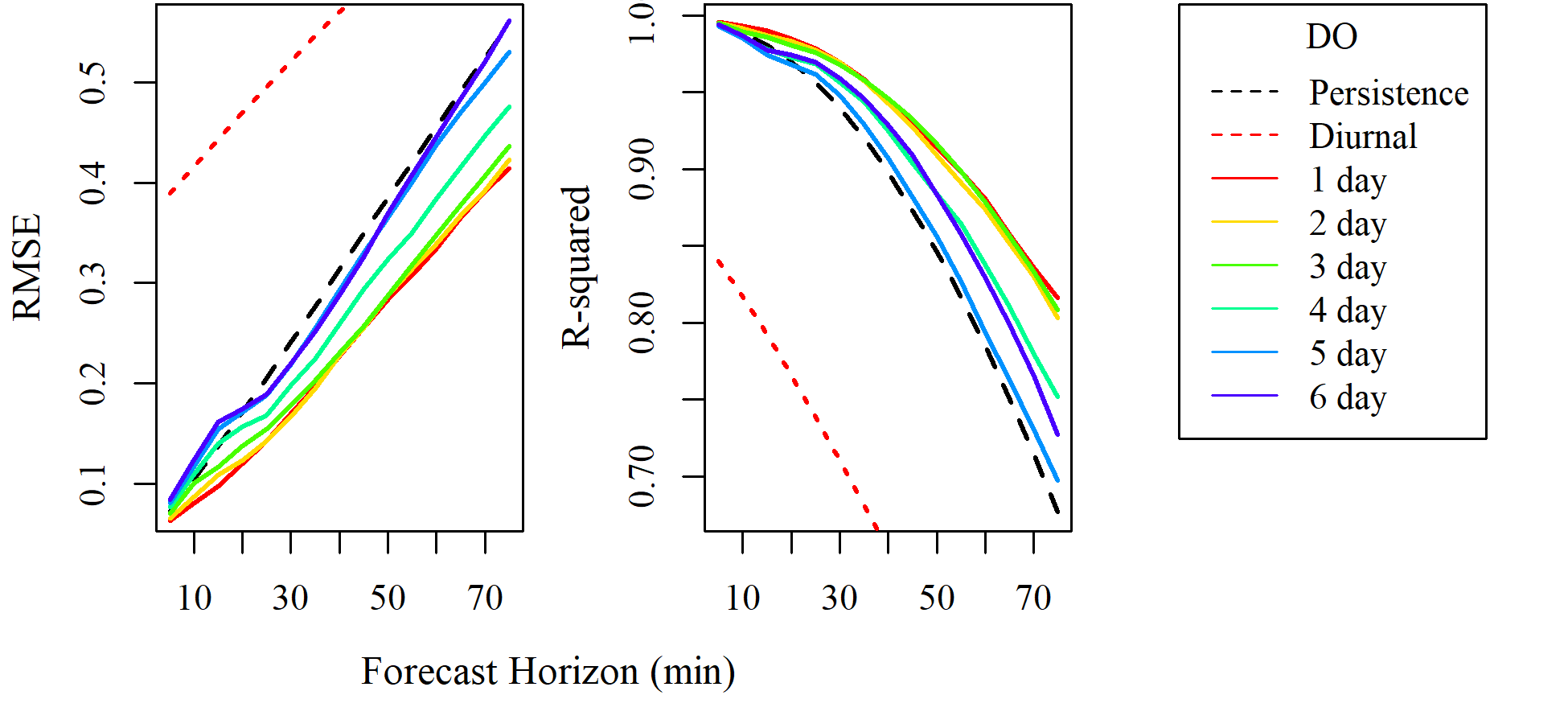
To accurately forecast ammonia, a multiple linear regression model is used that incorporates the time-dependent diurnal terms and monitored process variables. When the linear model is fit to various forecast horizons of ammonia, it is evident that the forecast horizon and the length of the training dataset are related (Figure 7). For short forecast horizons (i.e., 5 minutes), the longer training windows are better able to fit the training data. As the forecasting horizon increases, shorter training windows (i.e., 1 day) are the most accurate. A contributing factor may be the difference in the ammonia over weekends as opposed to weekdays, which tend to exibit stronger regular diurnal patterns.

NEEDS DONE –> Subtract off diurnal component to calculate fraction accounted for by process variables vs time NEEDS DONE –> Look are variables included in the model

 [Figure 7].{custom-style=“Figure-Label”}

Overlay ammonia?

An analysis of testing model error (i.e., accuracy of model from data not used in model training) conveys a different message than te error observed during training. Figure 8 is an example of the DO control and ABAC 4.0, 90s configuration (testing RMSE and R2 values for each training window, forecast horizon, and test configuration can be found in the [Appendix](#model-testing-error)). When only the diurnal model is considered (i.e., time as a function of sine and cosine are the only predictors), the existing control paradigm of using the current measured sensor value of ammonia (i.e., *persistence* forecast) is more accurate (i.e., lower RMSE and higher R2). When the linear model is trained using time and process variables, the combination diurnal-linear model is more accurate than the current control paradigm.

 Figure 8. Testing RMSE and R-squared for the models tested on the DO control and ABAC 4.0, 90s confiuration. The black dashed line is the error of the *persistence* forecast, which simulates the current operating condition of using the ammonia sensor’s current measured value to respond to water quality changes. The red dashed line is the diurnal model error trained on 1 day of data. The 2-day diurnal model is within the plotting region for ABAC 4.0, 90s, but all other diurnal model training windows performed too poorly to be visualized.

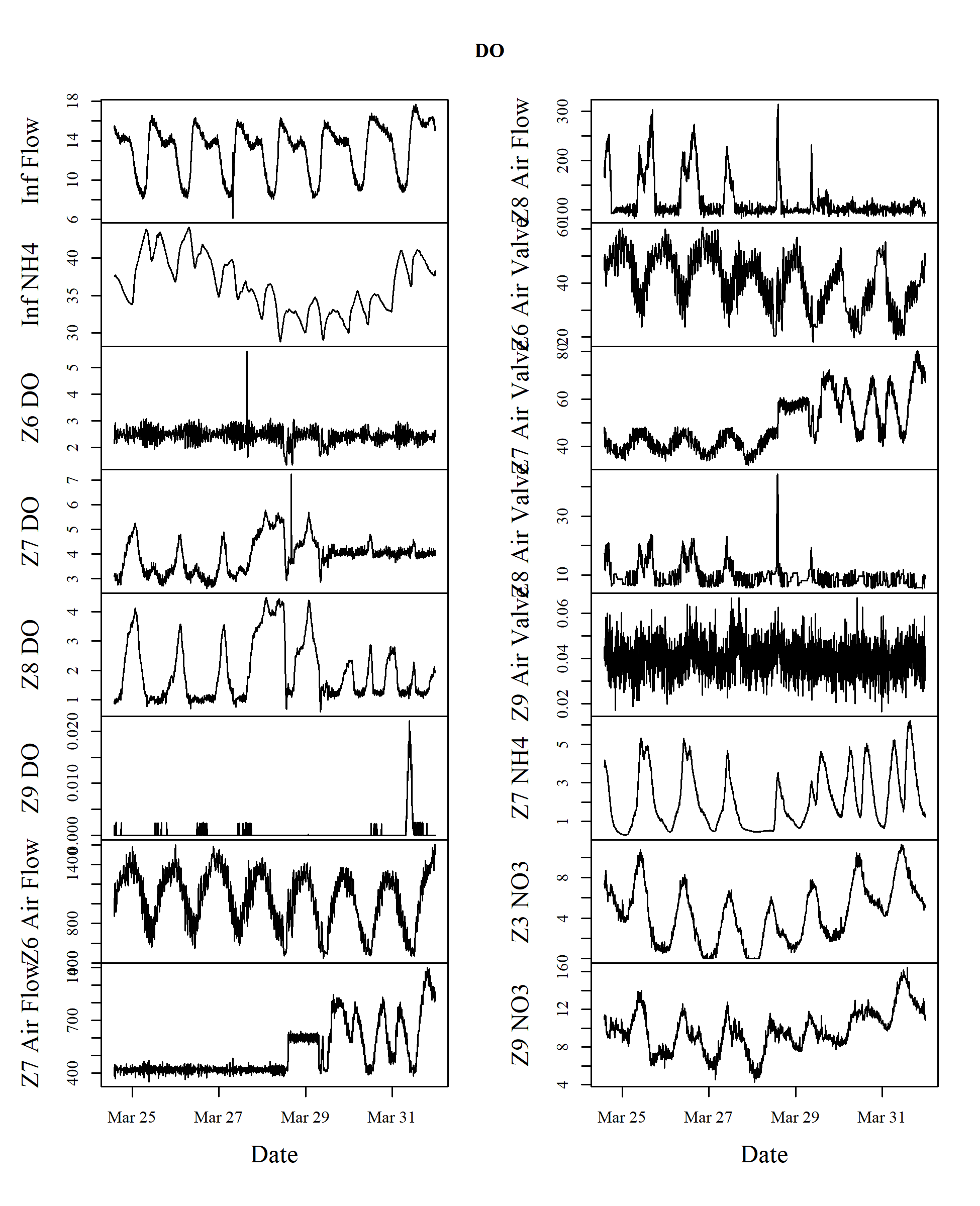
To demonstrate the potential improvement by incorporating the forecasting model into the cascade control loop, the percent improvement between the persistence forecast and the model forecast for each training window and forecast horizon are calculated and averaged across the four control conditions evaluated in this study (Table 2). Depending on the desired forecast horizon, a training window can be selected that has the largest percent improvement (i.e., reduction in error or RMSE)

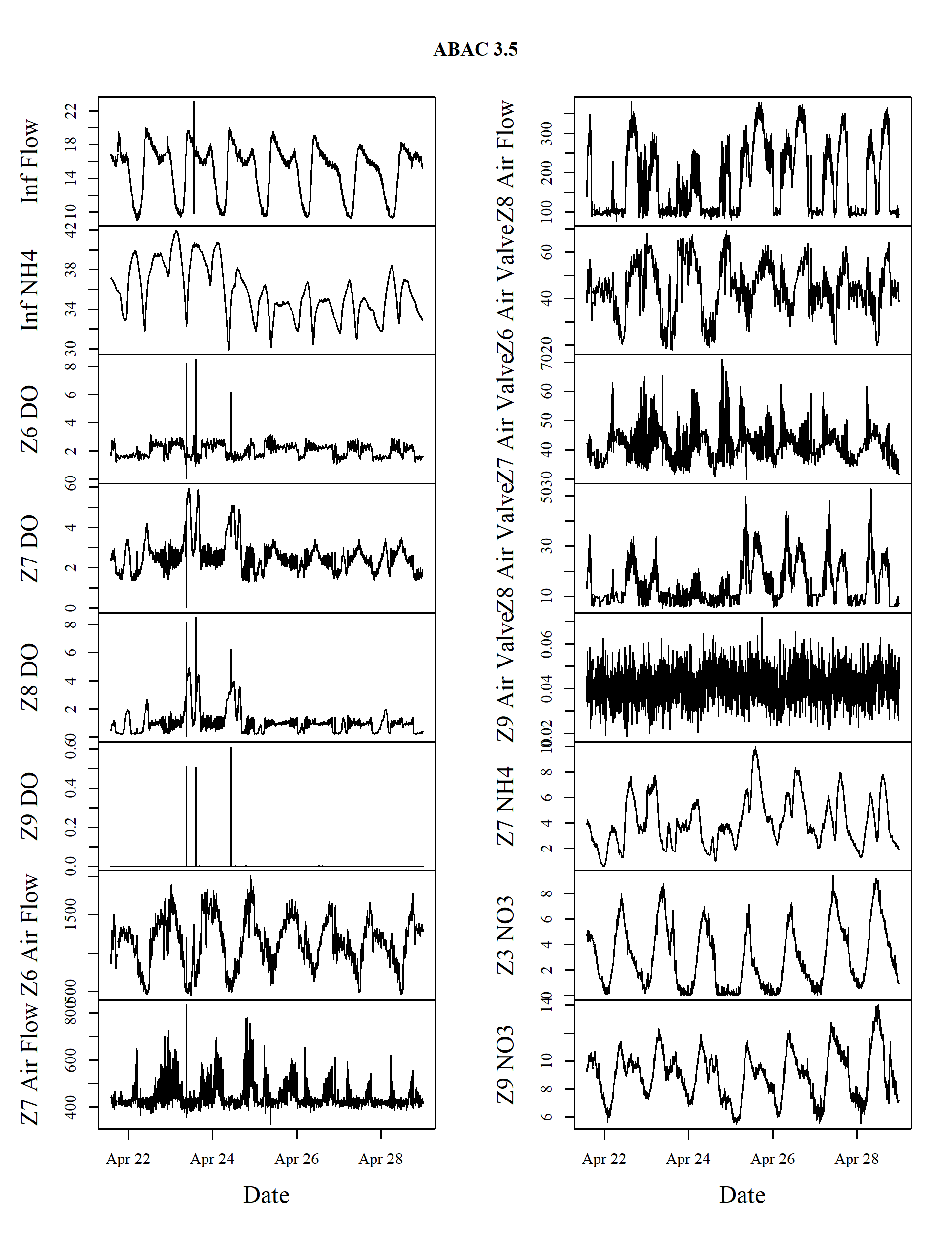
Table 2. Average percent improvement for all control configurations (DO and ABAC) by training window length and forecast horizon. Positive values indicate the model forecast was more accurate than the persistence forecast.

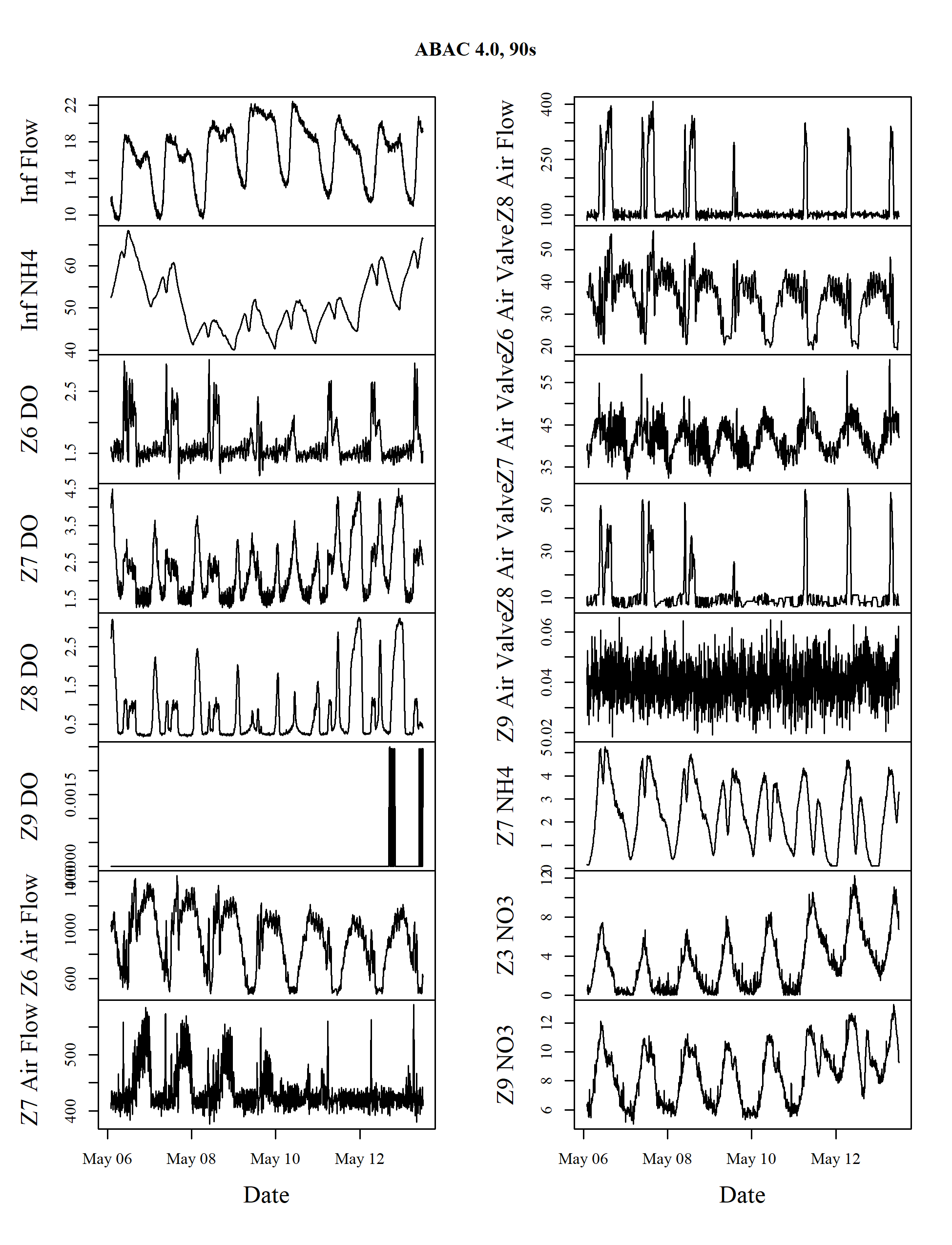
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 5 Min | 15 Min | 25 Min | 35 Min | 45 Min | 55 Min | 65 Min | 75 Min |
| 1 Day | 1.6 | 11 | 7.2 | 4.2 | 2.7 | 3.3 | 3.4 | 3.2 |
| 2 Day | -2.3 | 16 | 17 | 12 | 6.9 | -2.9 | -7.8 | -14 |
| 3 Day | -3.8 | 12 | 20 | 21 | 21 | 20 | 20 | 20 |
| 4 Day | -2.9 | 11 | 25 | 25 | 24 | 23 | 23 | 24 |
| 5 Day | -3 | 7.5 | 25 | 25 | 25 | 25 | 25 | 26 |
| 6 Day | -2.7 | 6.8 | 28 | 30 | 31 | 31 | 31 | 31 |

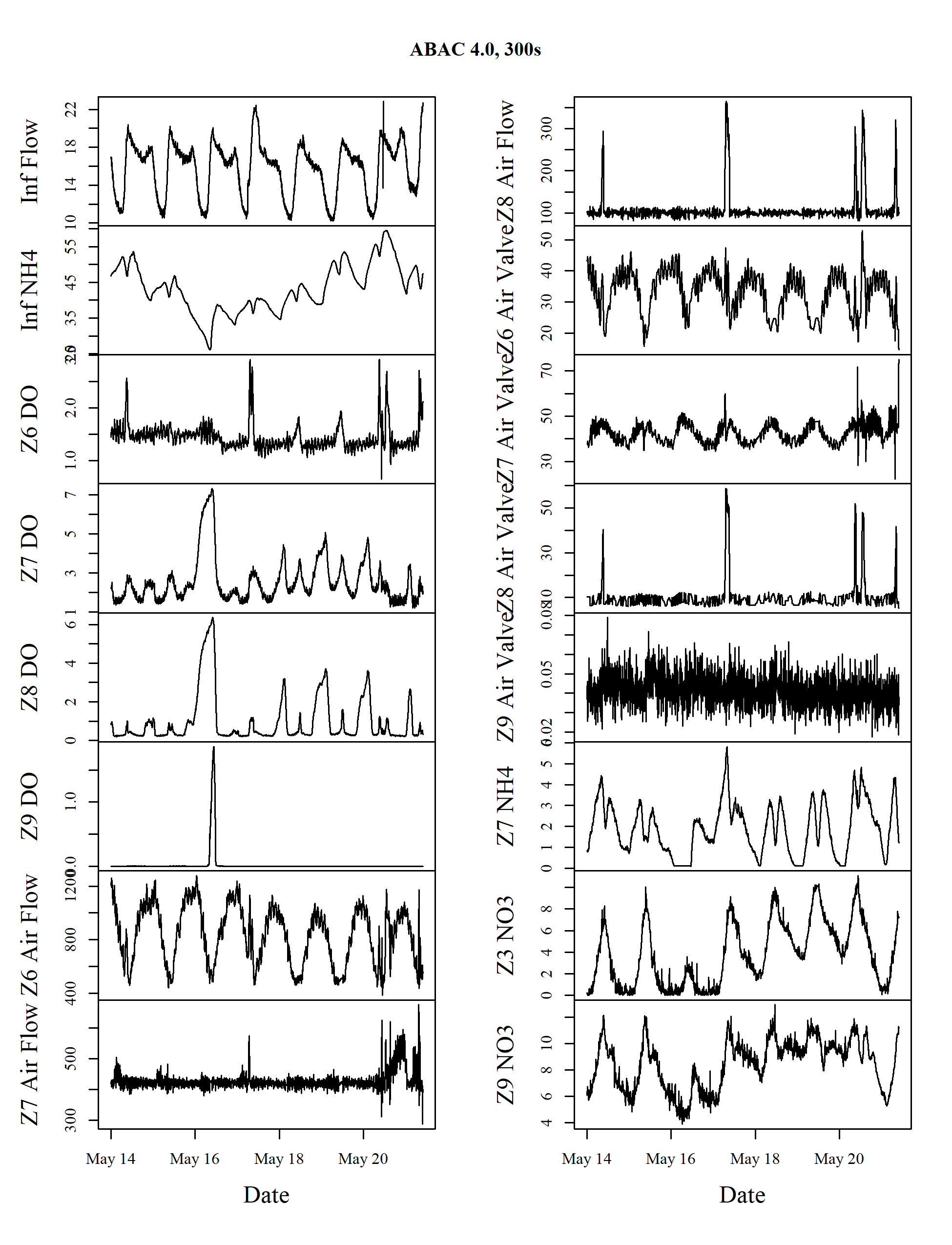
# Appendix

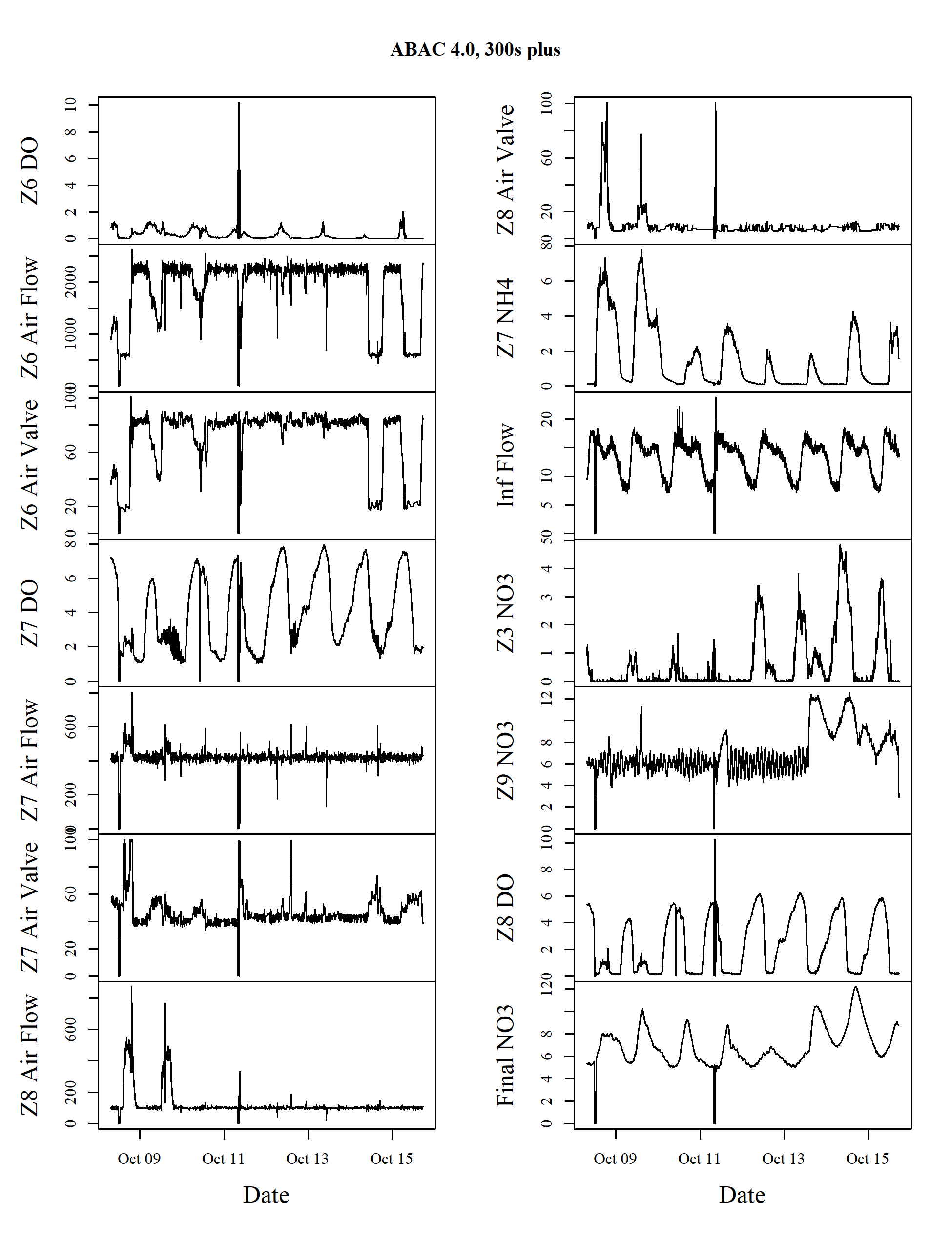
## Timeseries

 Figure 1. Timeseries plots of all data for the DO control configuration

 Figure 2. Timeseries plots of all data for the ABAC 3.5 control configuration

 Figure 3. Timeseries plots of all data for the ABAC 4.0, 90s control configuration

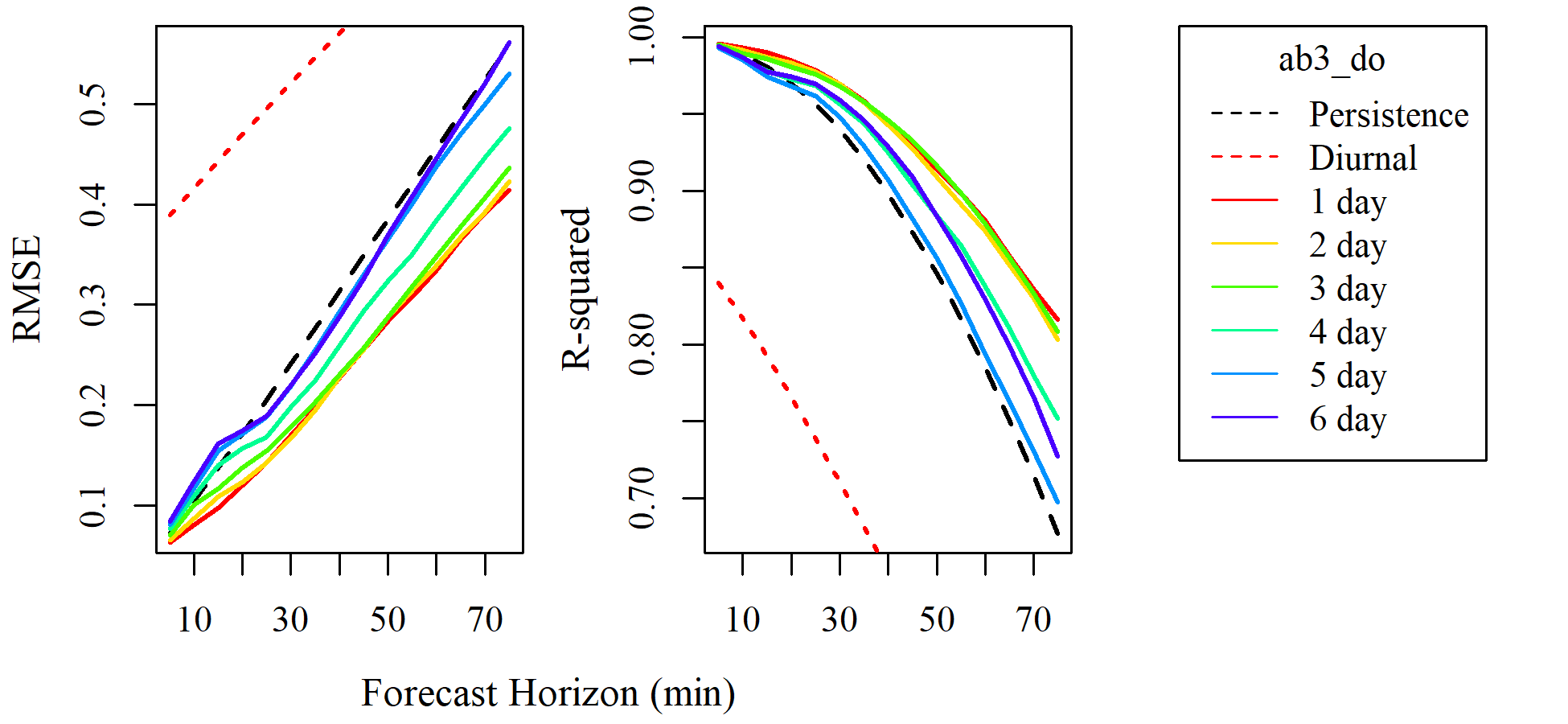
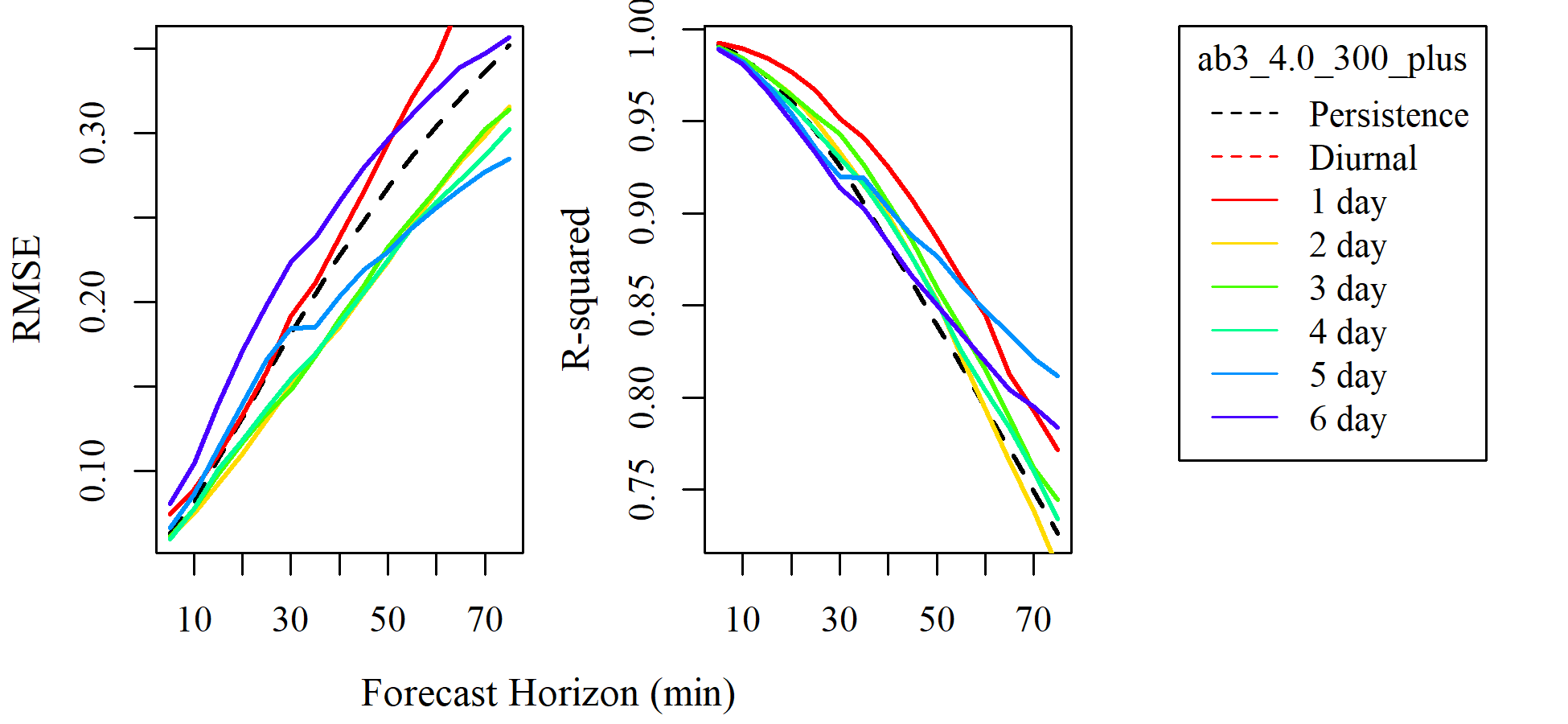
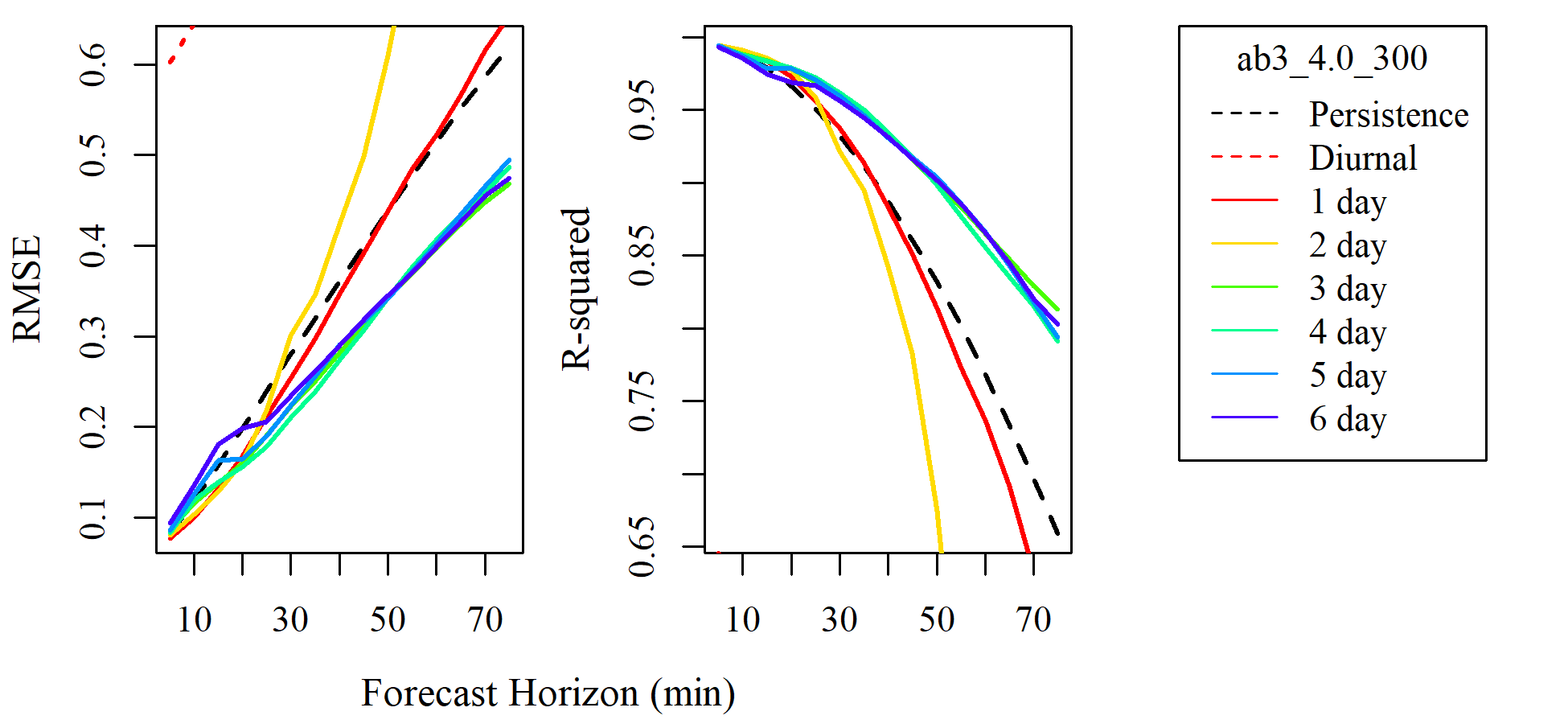
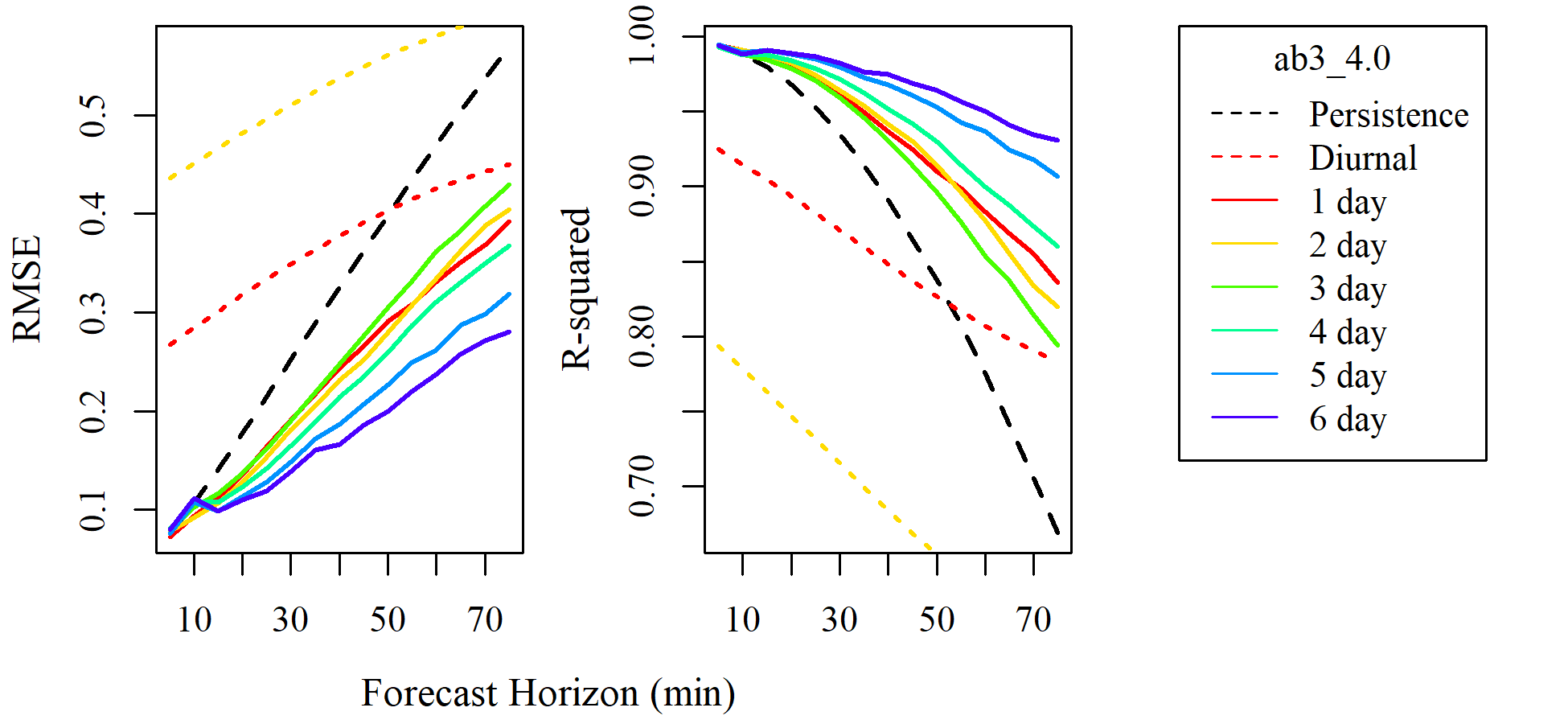
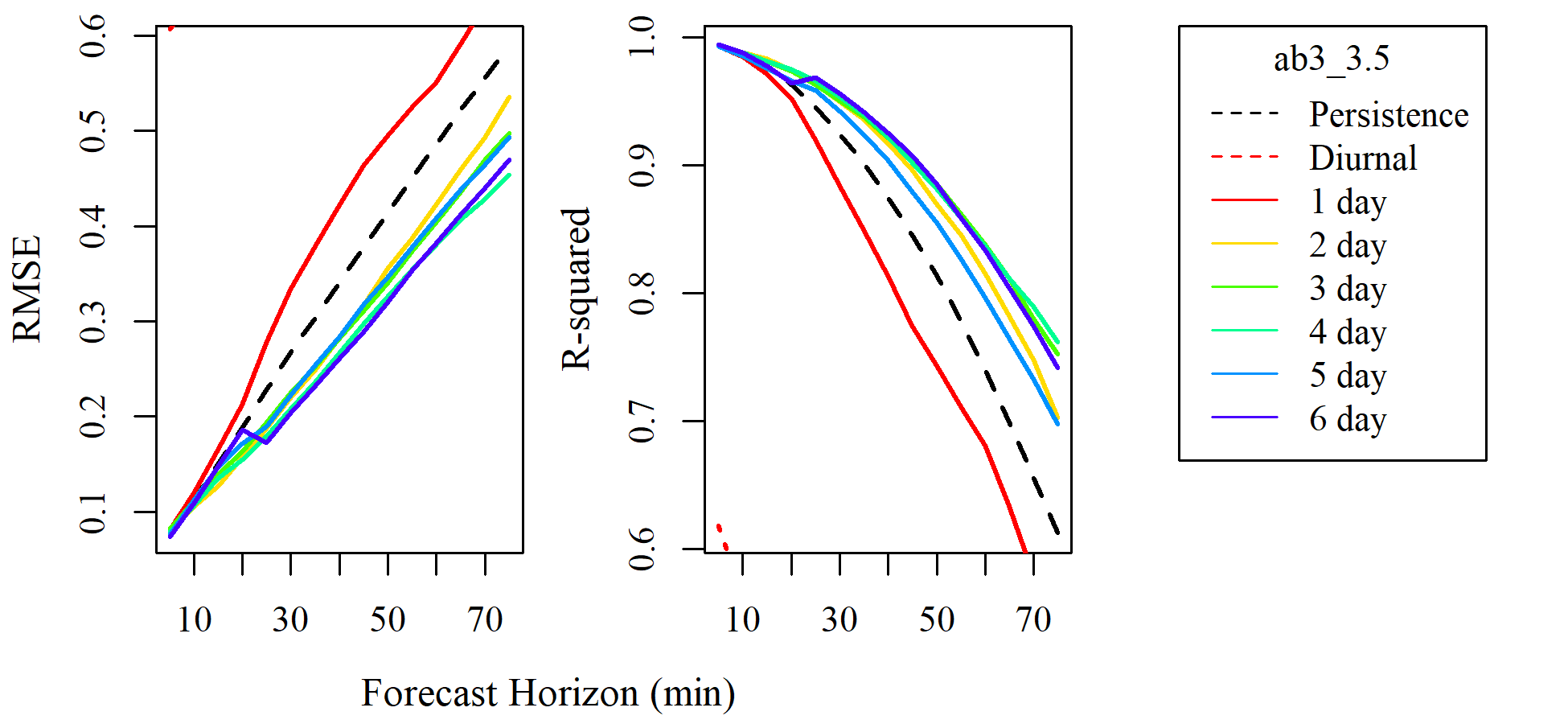
 Figure 4. Timeseries plots of all data for the ABAC 4.0, 300s control configuration

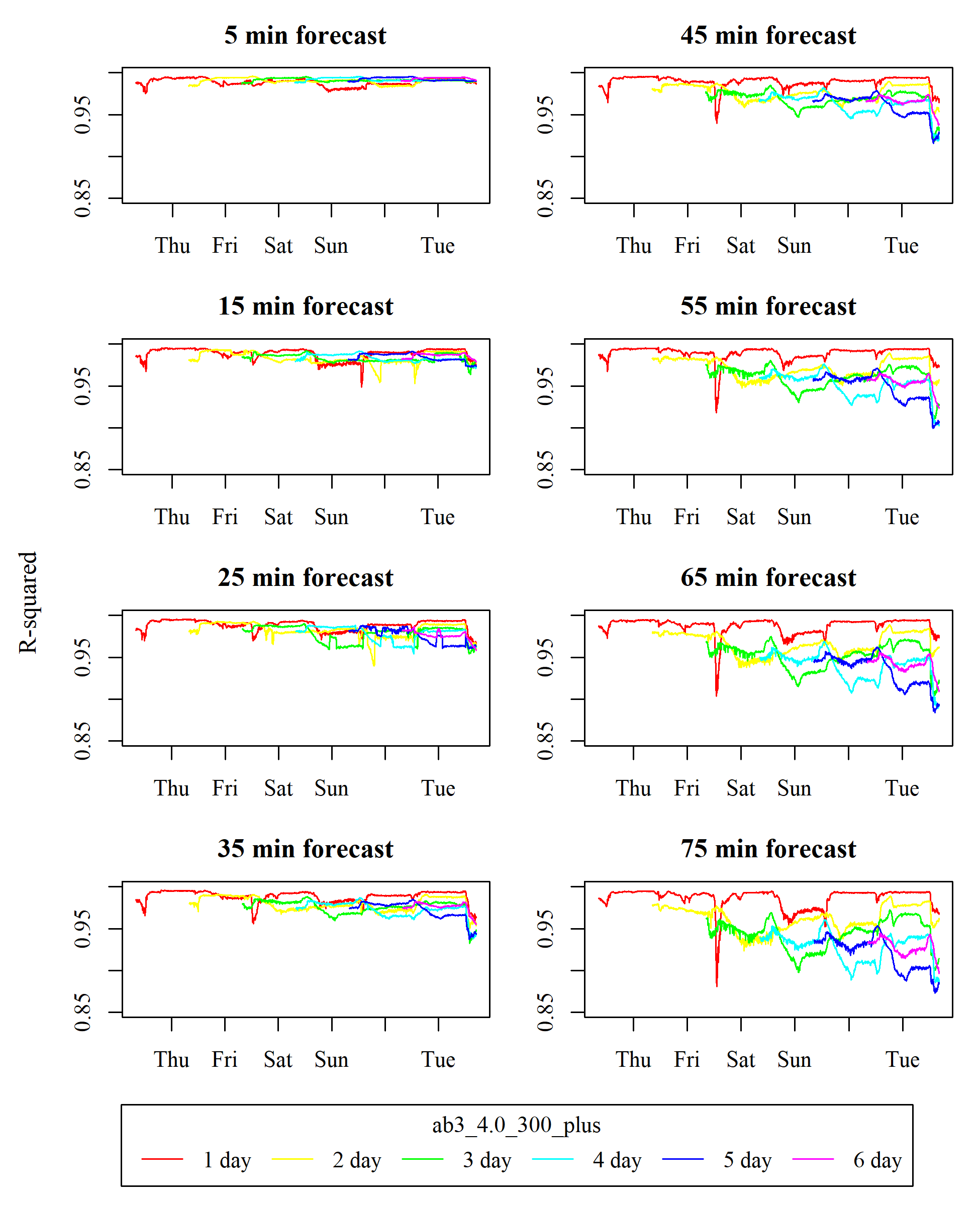
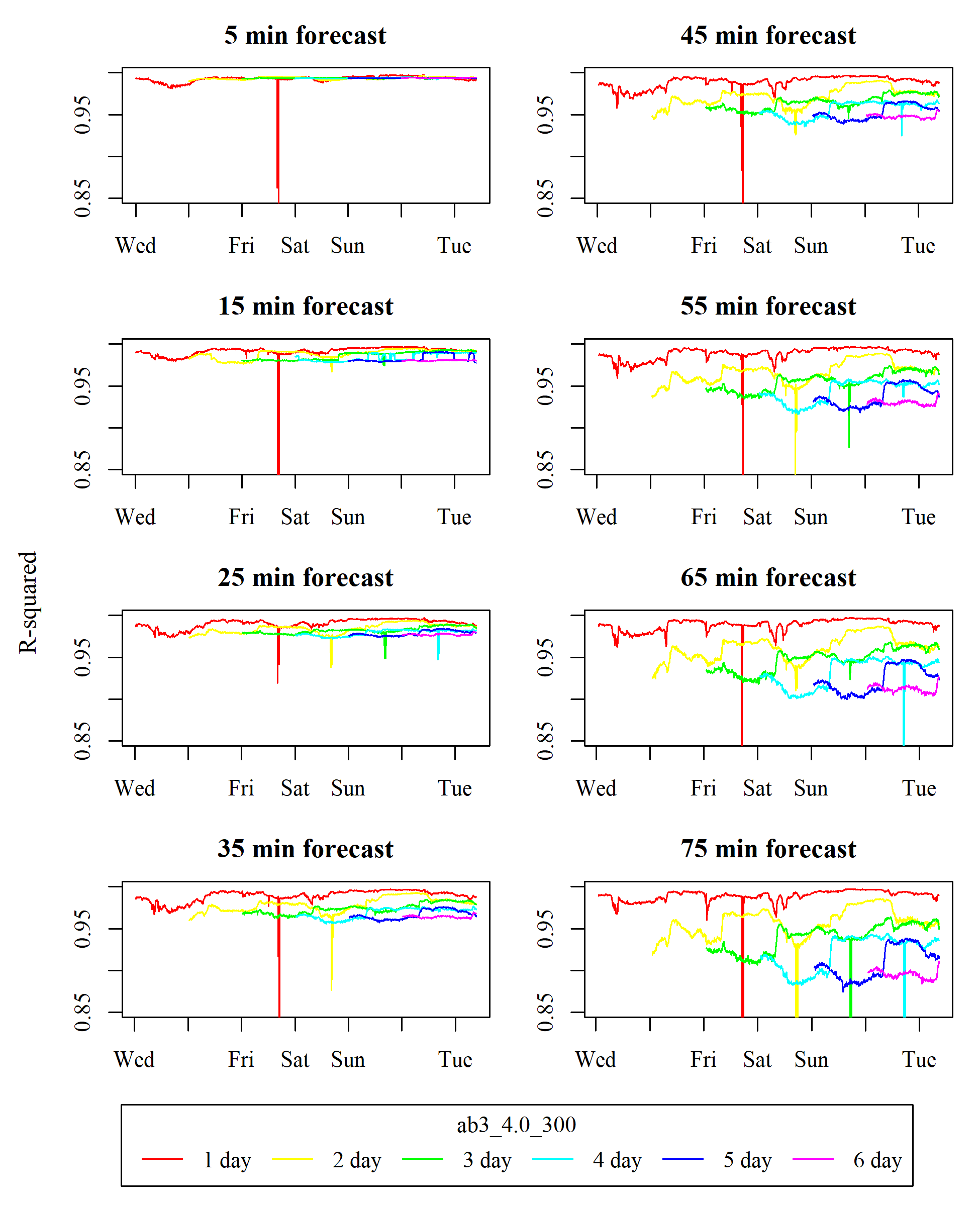
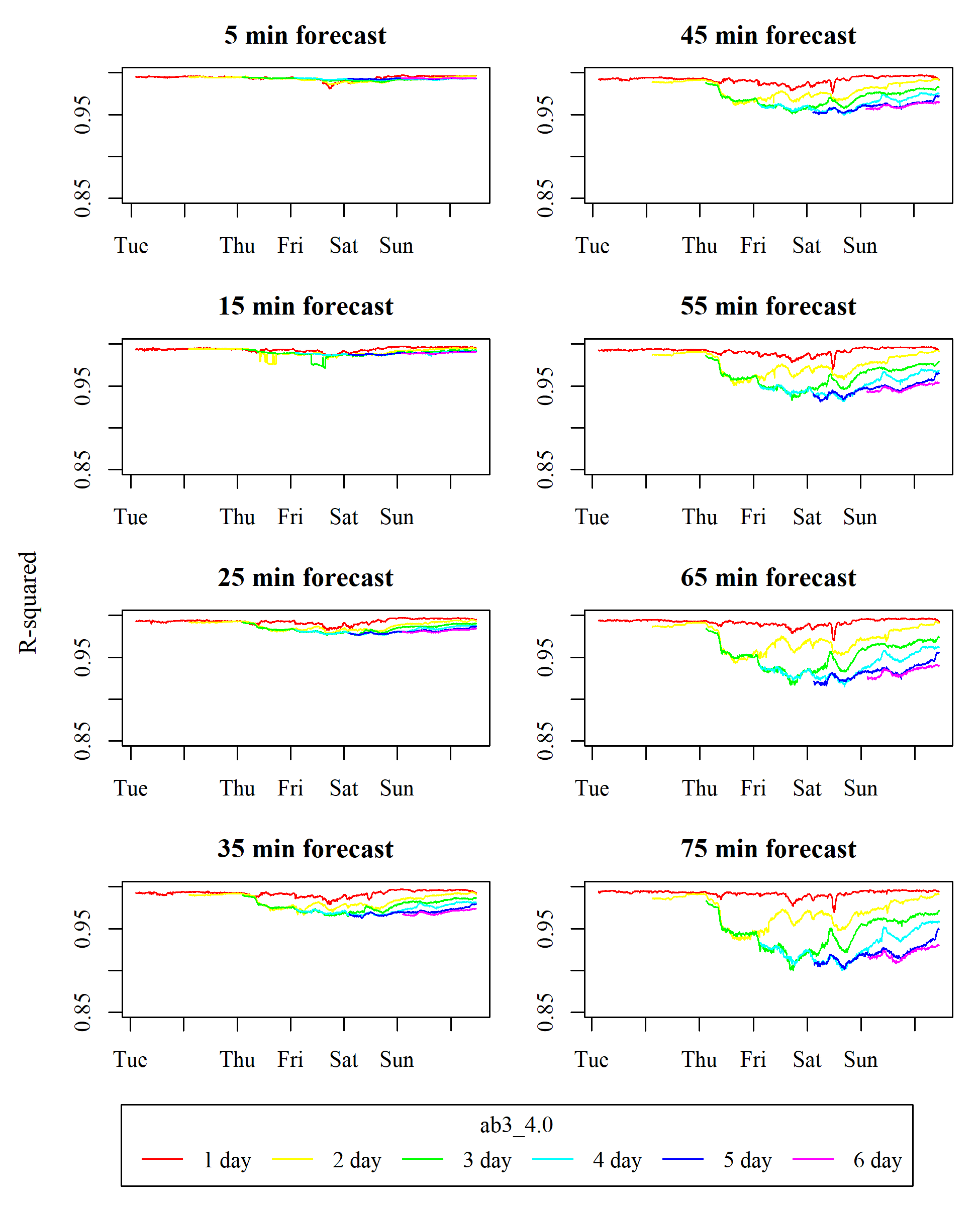
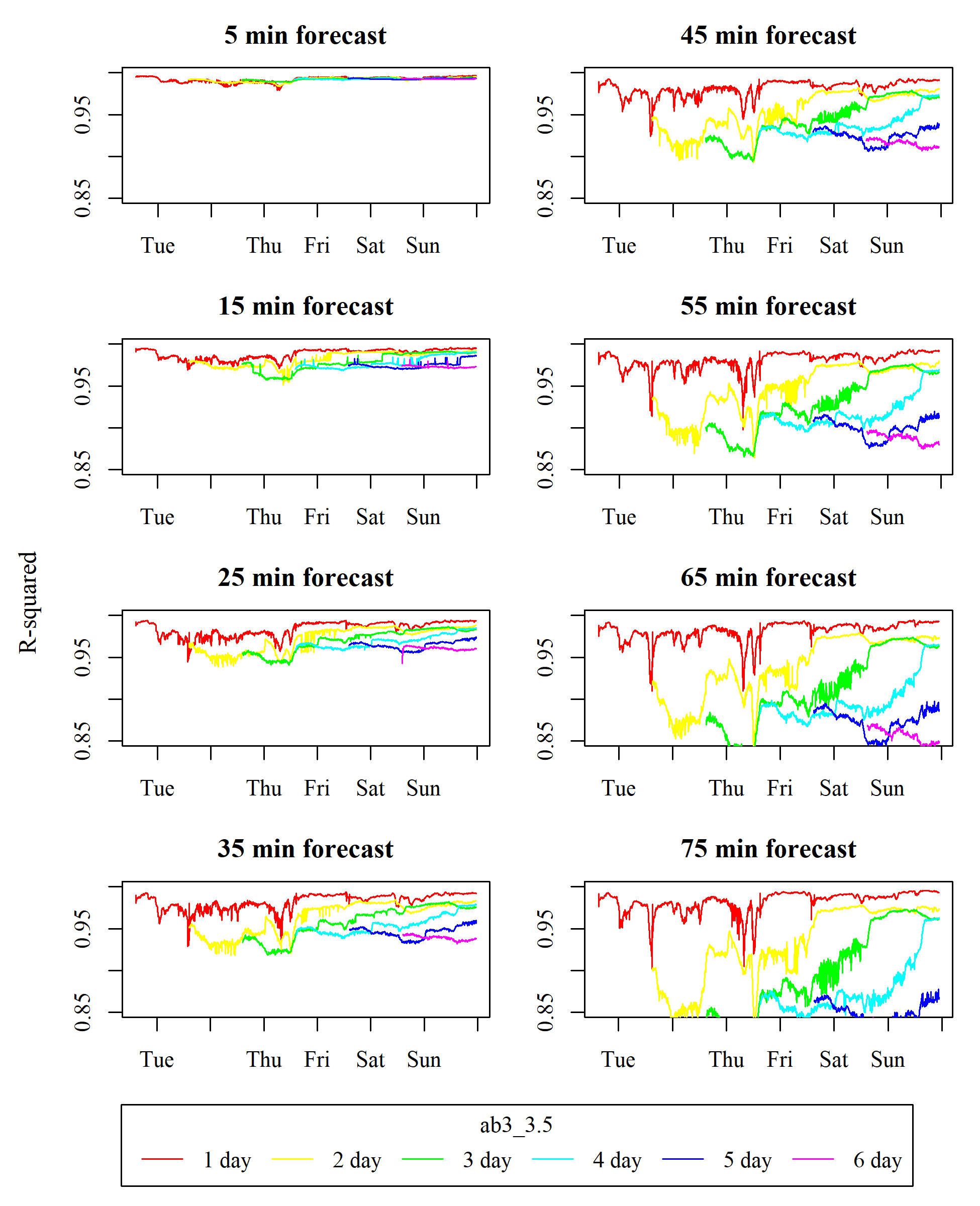
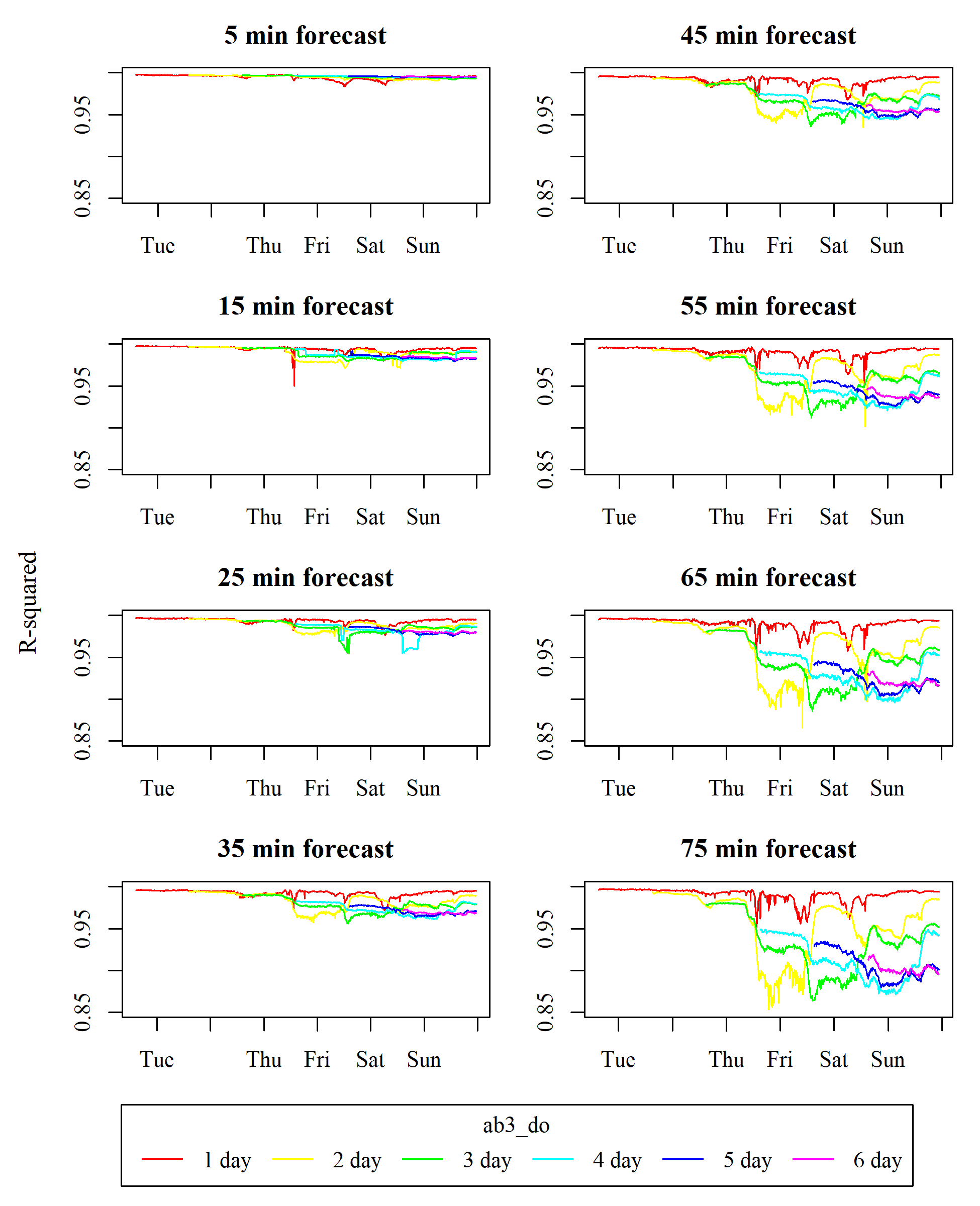
 Figure 5. Timeseries plots of all data for the ABAC 4.0, 300s plus control configuration

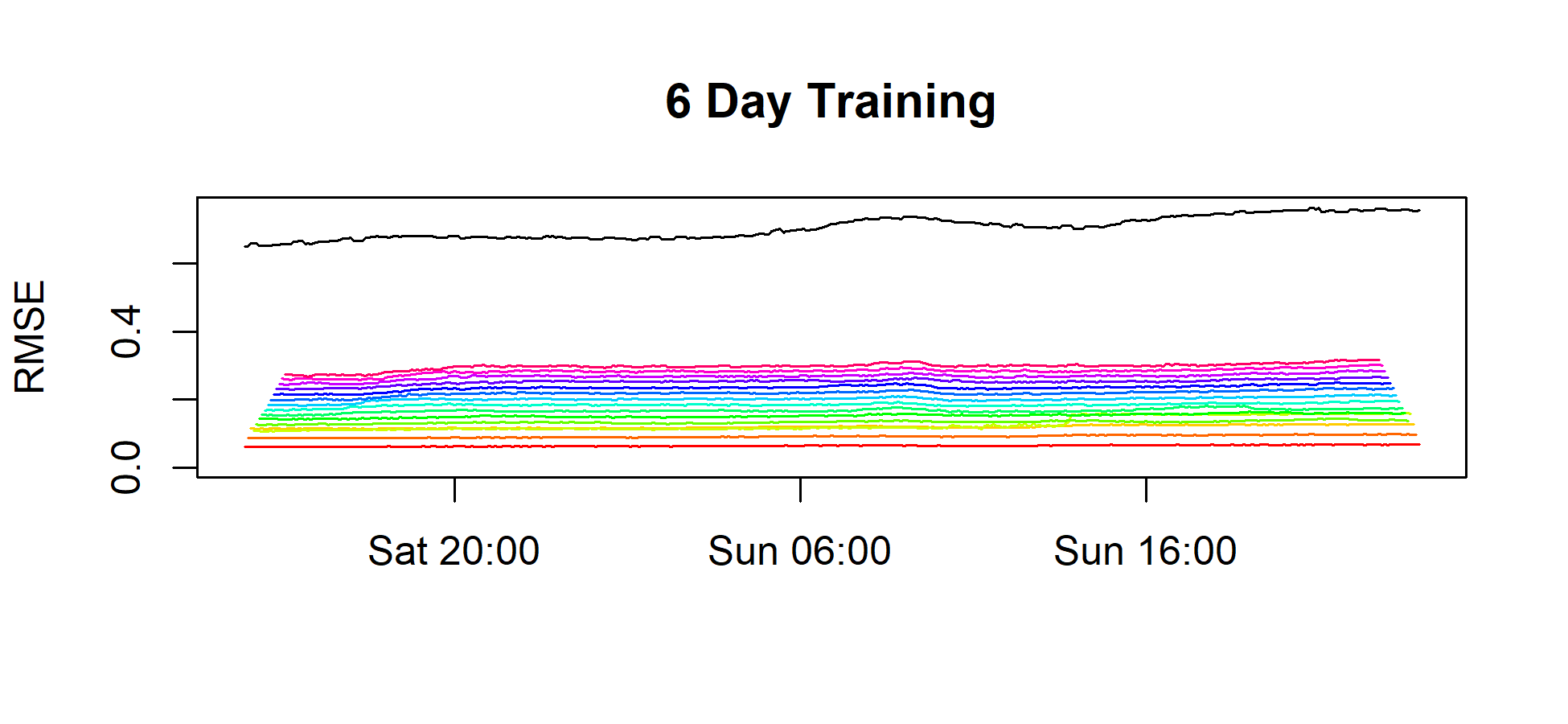
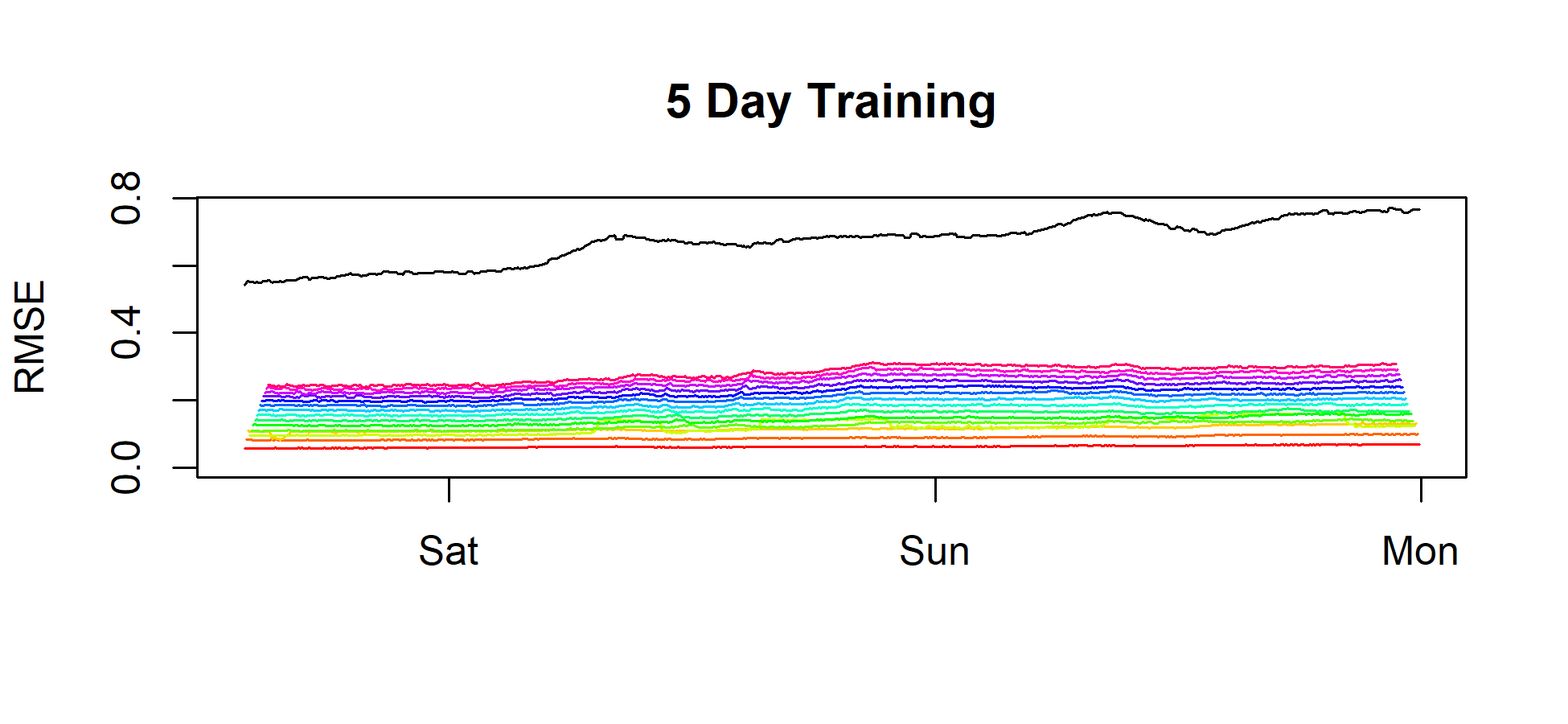
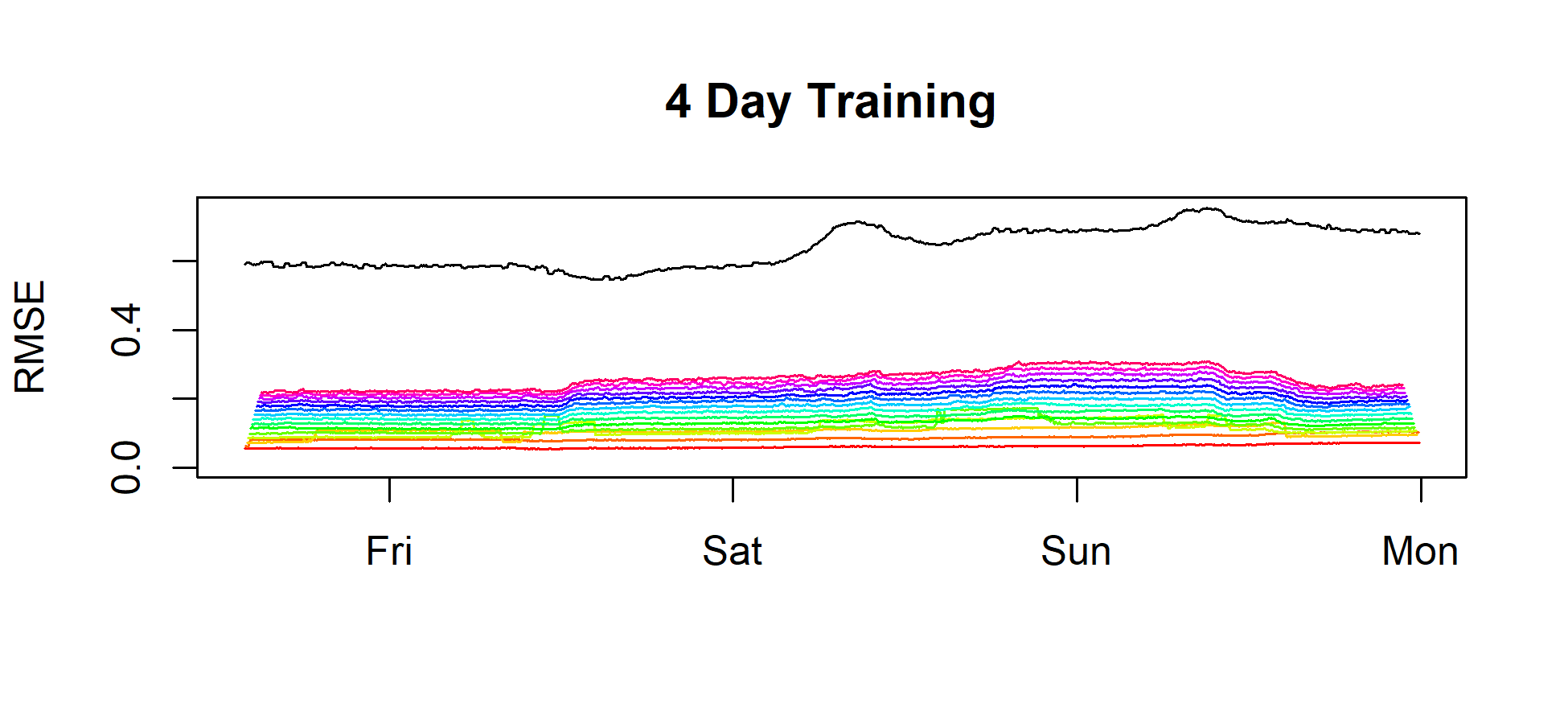
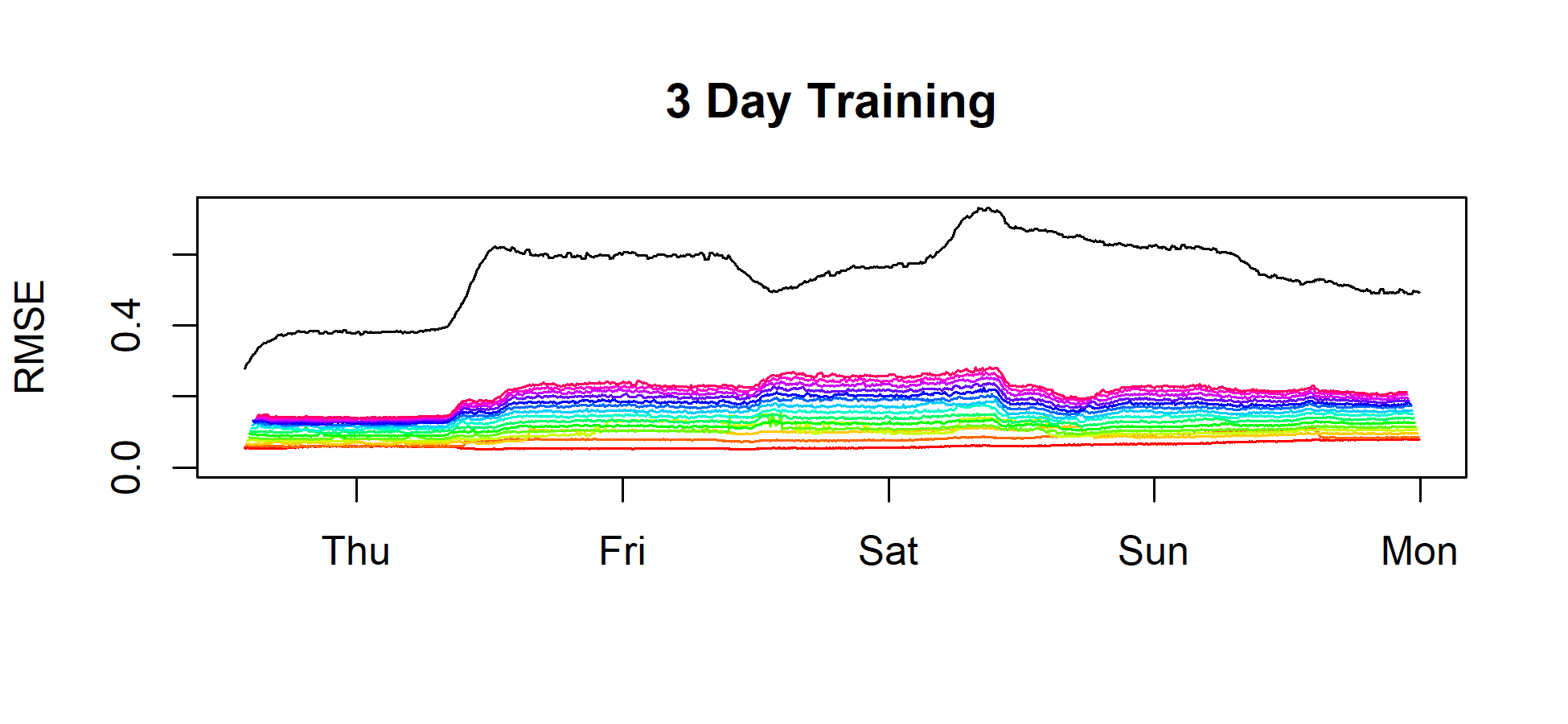
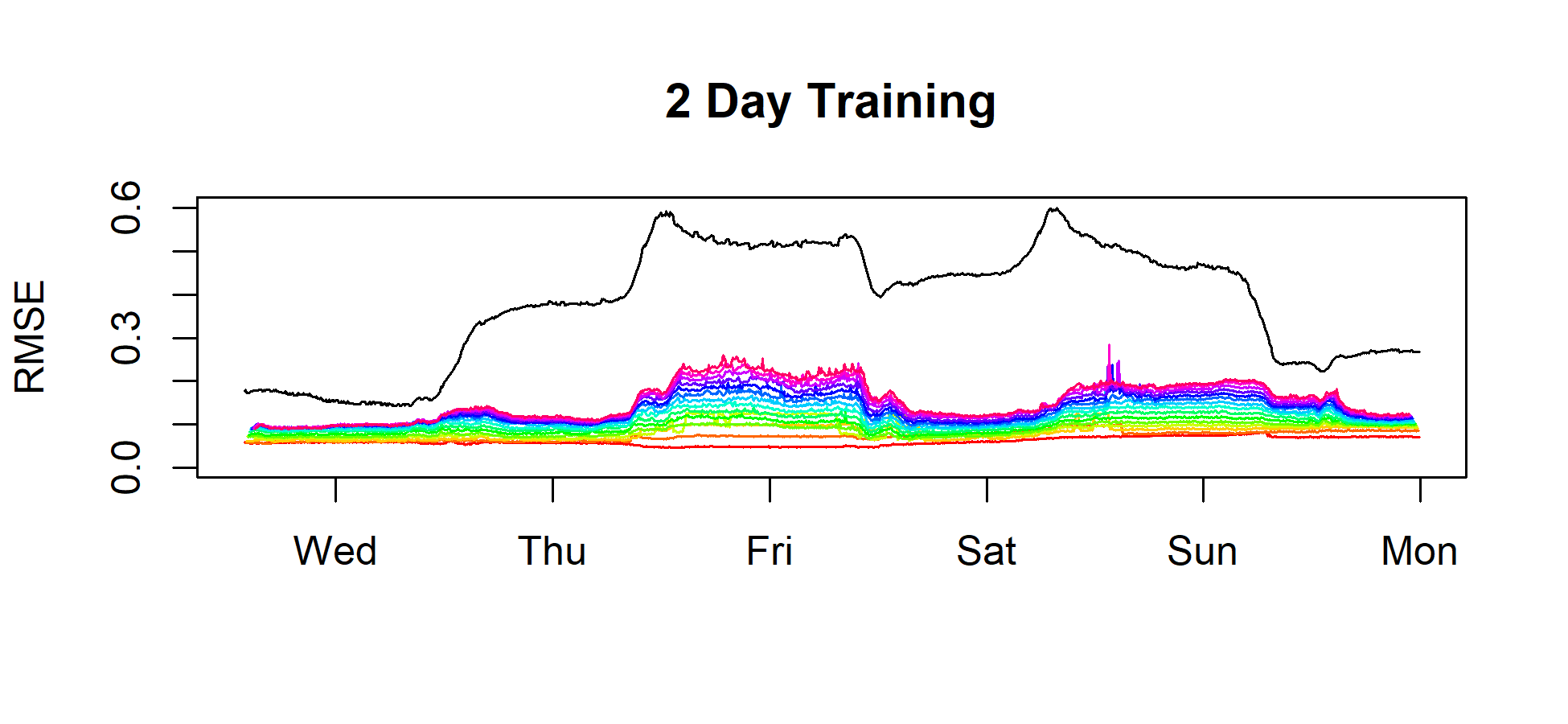
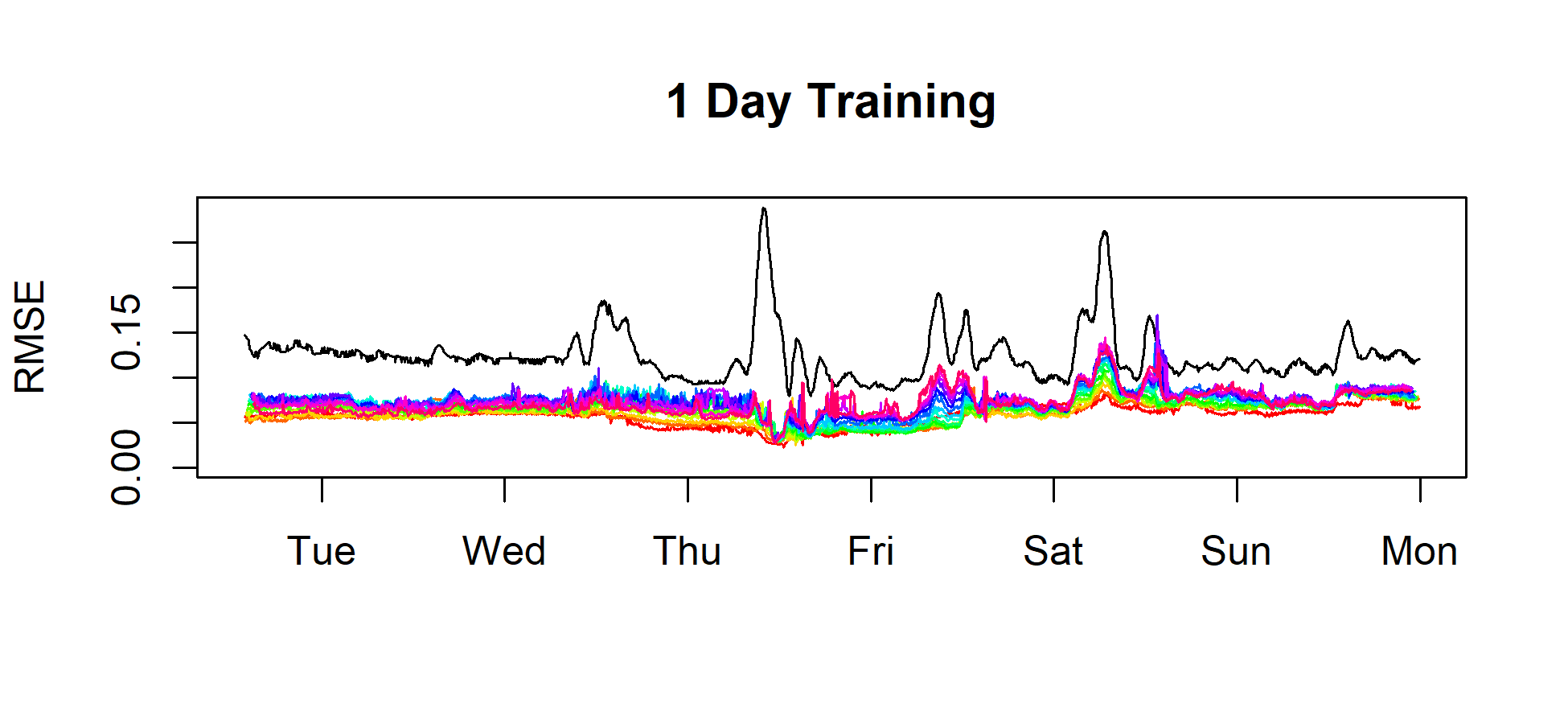
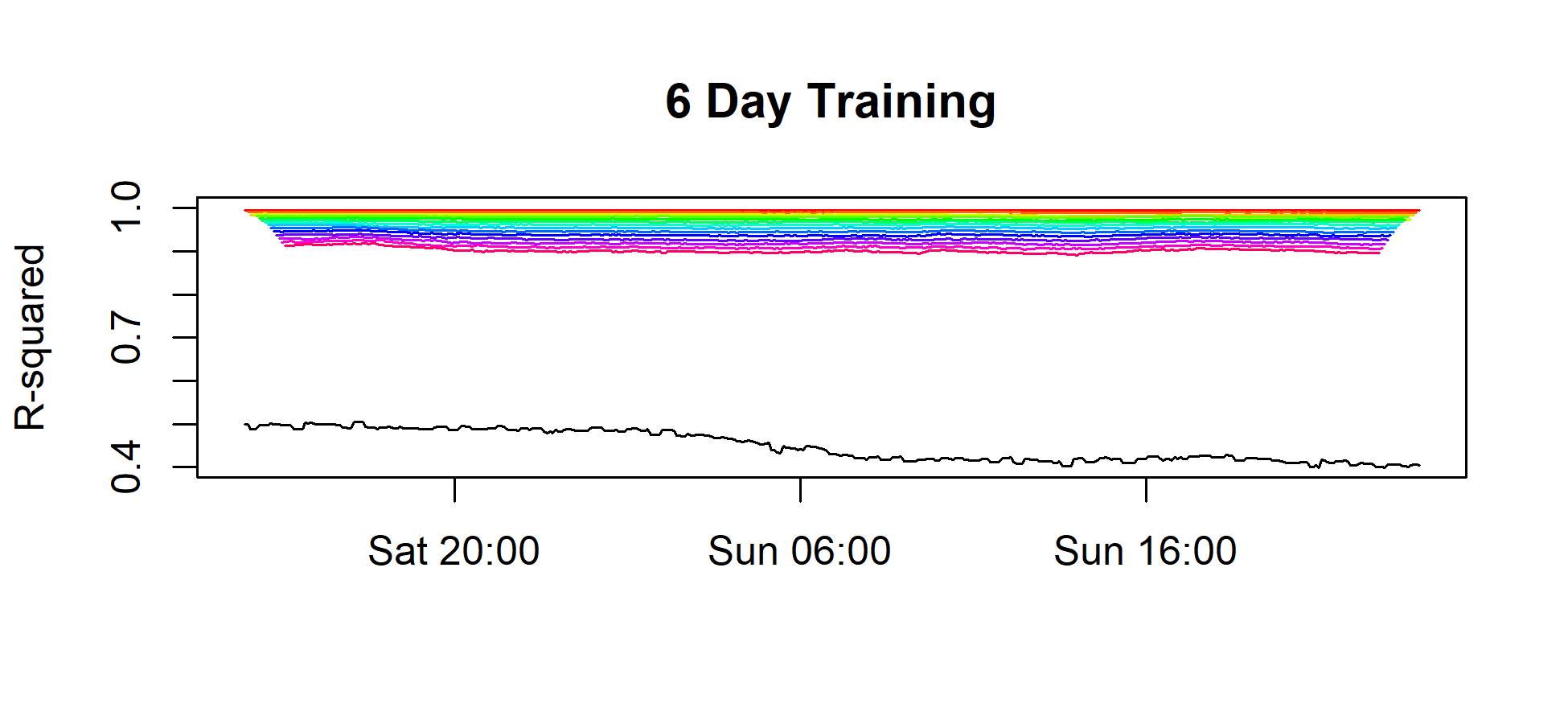
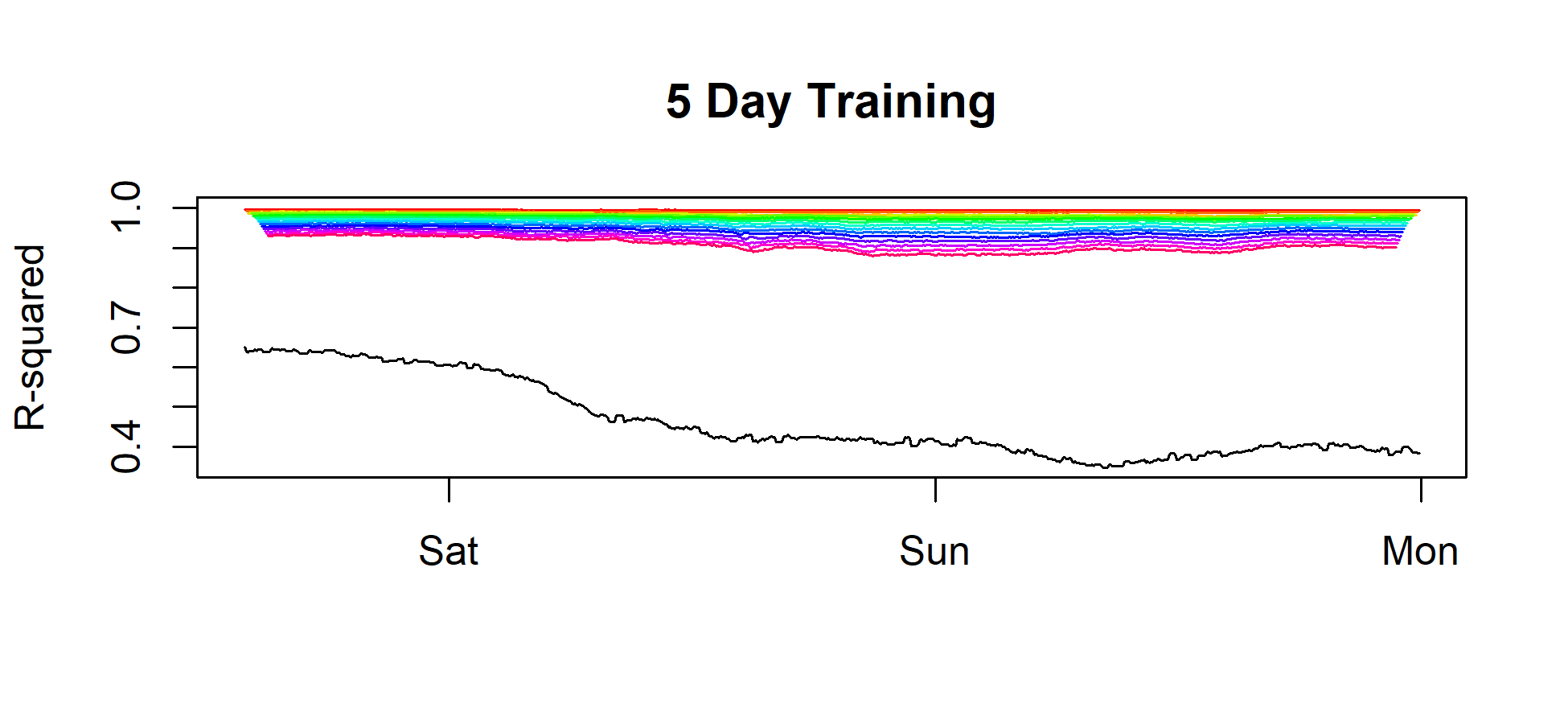
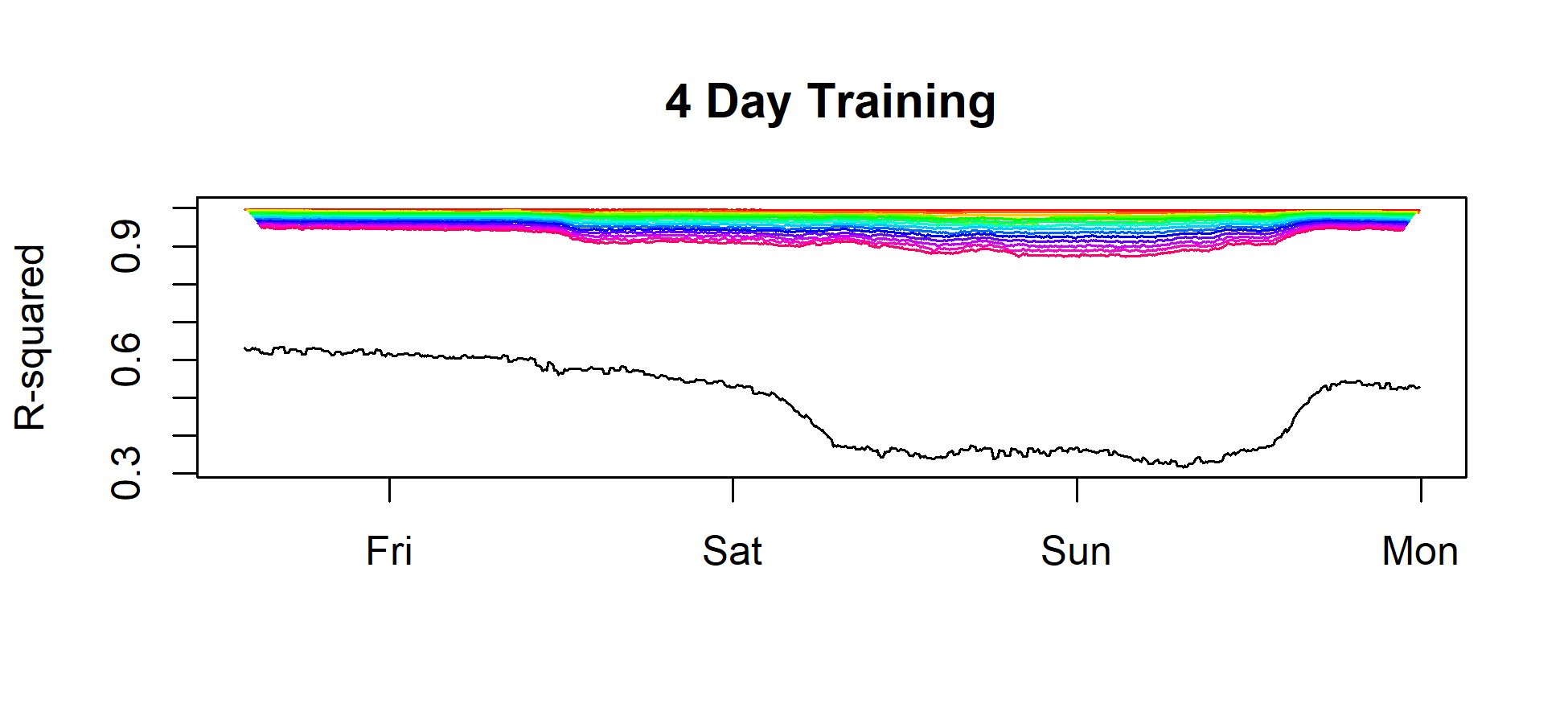
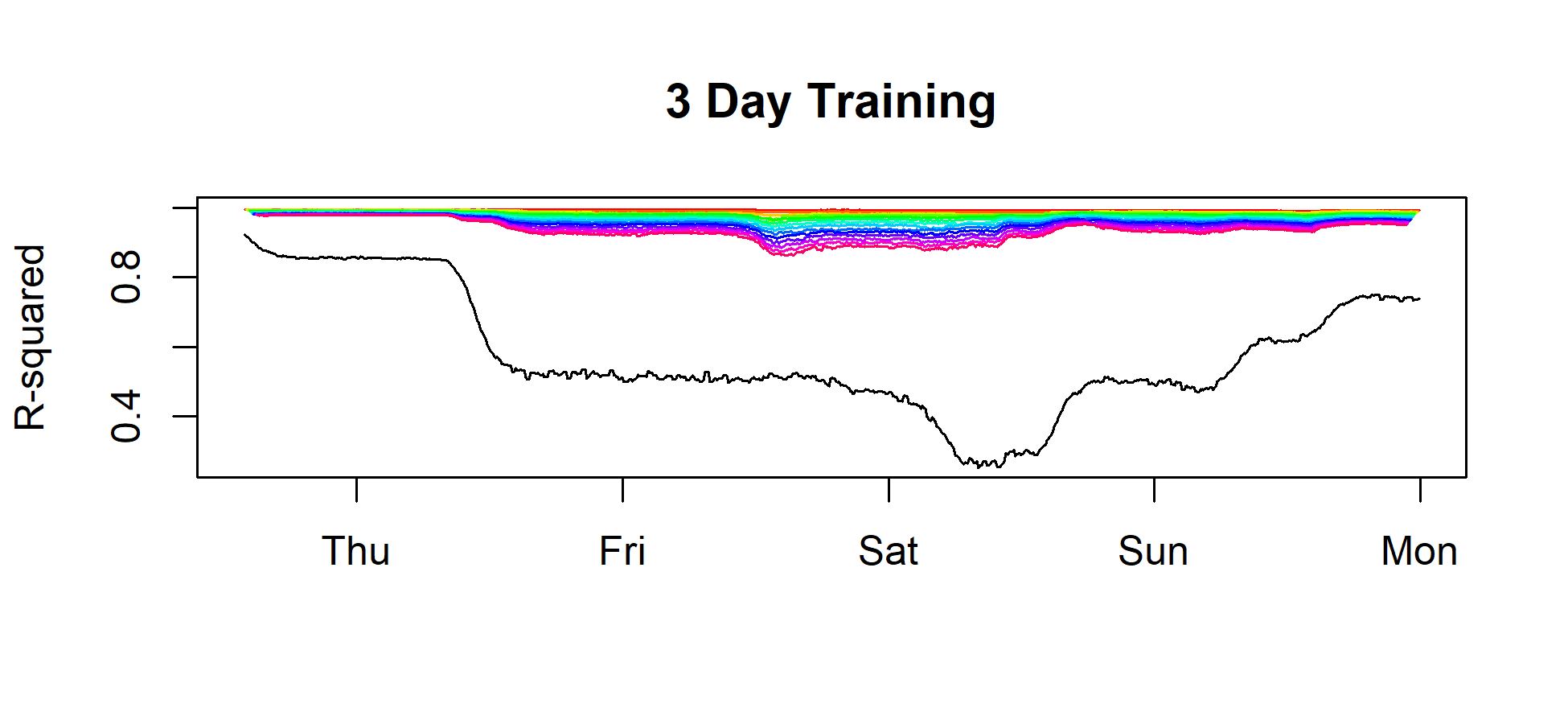
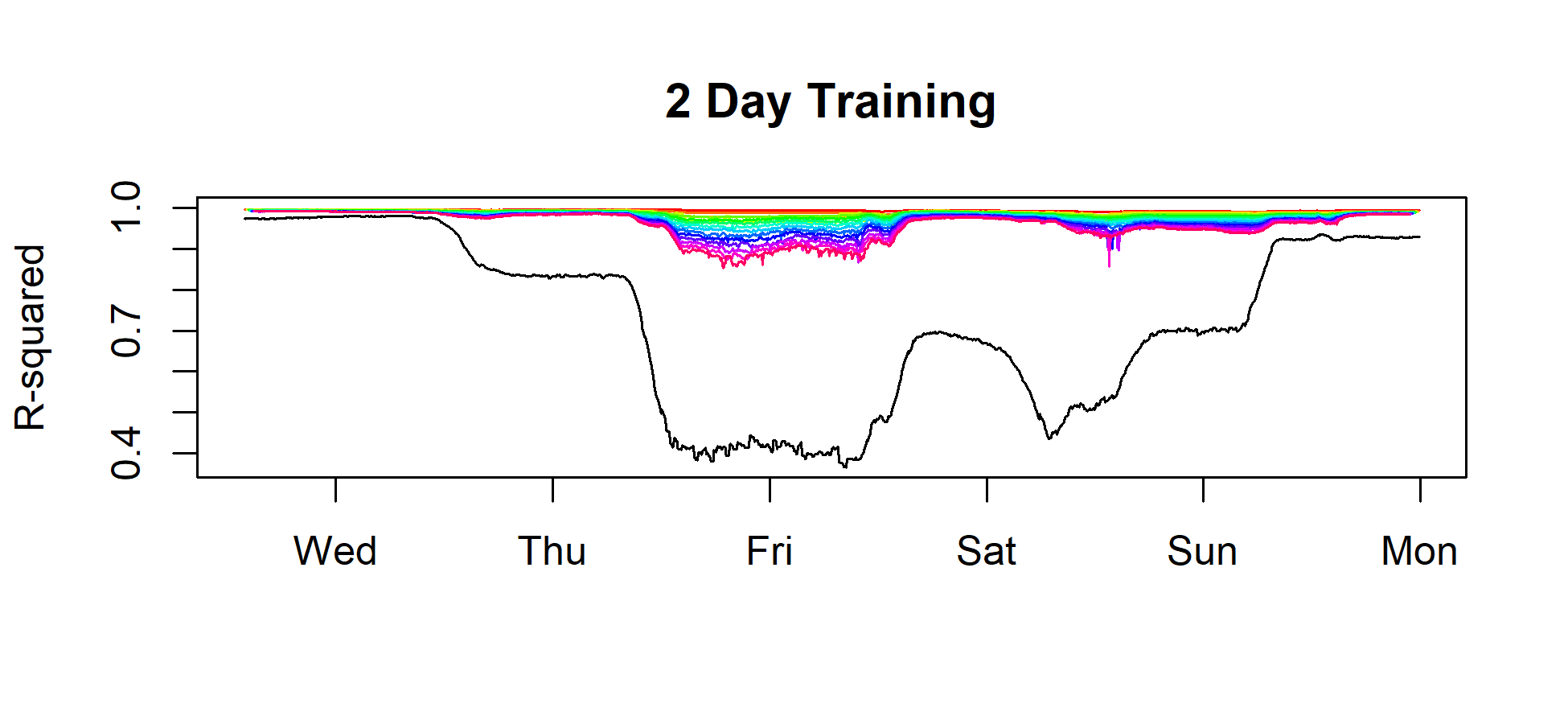
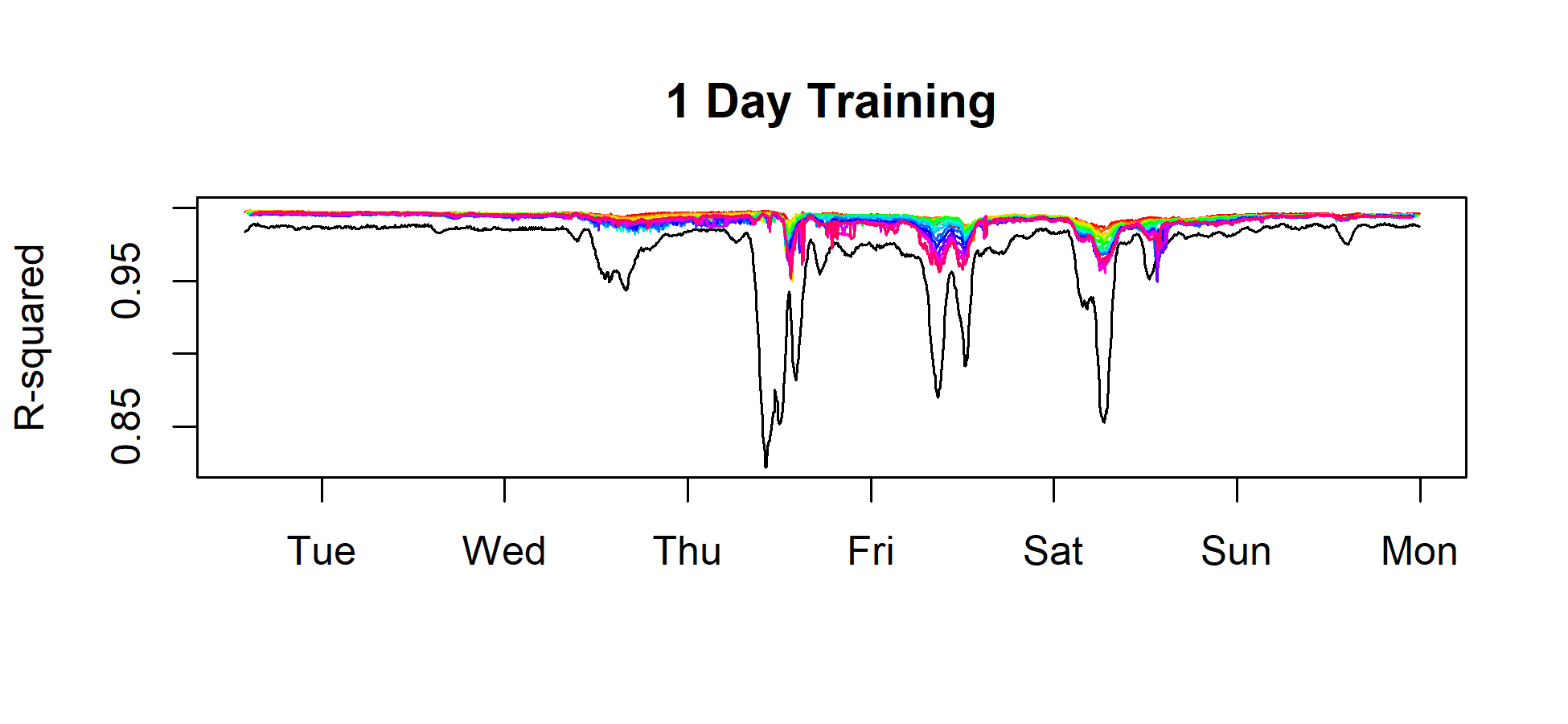
## Model training error

library(xts)  
# Just DO for now  
# all.days.files <- load("results/all-days-ab3\_do.RData", envir=.GlobalEnv)  
# all.days.files <- sapply(list.files("results/",pattern="all-days-"), function(x) load(paste0("results/",x), envir=.GlobalEnv))  
sapply(list.files("results/",pattern="all-days-"), function(x) load(paste0("results/",x), envir=.GlobalEnv))  
all.days.files <- c("all.days.ab3\_do",  
 "all.days.ab3\_3.5",  
 "all.days.ab3\_4.0",  
 "all.days.ab3\_4.0\_300",  
 "all.days.ab3\_4.0\_300\_plus")  
  
for(file in all.days.files) {  
 all.days <- get(file)  
 all.horizons.training.error <- list()  
 for(step in 1:length(all.days[[1]])) {  
 for(days in 1:length(all.days)) {  
 obj.data <- all.days[[days]]  
 results <- obj.data[[step]]  
 temp.results <- xts(cbind(results$Training.R2.D,results$Training.R2.LM), order.by = results$Test.time)  
 colnames(temp.results) <- paste0(days,"Day.",c("Training.R2.D", "Training.R2.LM"))  
 if(days==1) {  
 r2.results <- temp.results  
 } else {  
 r2.results <- cbind(r2.results, temp.results)  
 }  
 }  
 all.horizons.training.error[[length(all.horizons.training.error)+1]] <- r2.results  
}  
  
layout.matrix <- matrix(c(1,2,3,4,9,5,6,7,8,9), nrow=5, ncol=2)  
layout(layout.matrix, widths=c(.5, .5), heights = c(0.225,0.225,0.225,0.225, 0.1))  
par(mar=c(2.5,2.5,2.5,1), oma=c(0,2,0,0), cex=.9, family="serif")  
for(step in 1:length(all.horizons.training.error)) {  
 if((step %% 2)==0) next  
 d.plot <- grep("D", colnames(all.horizons.training.error[[step]]))  
 lm.plot <- grep("LM", colnames(all.horizons.training.error[[step]]))  
 plot.colors <- rainbow(length(lm.plot))  
 # Plot diurnal training r2 for each training window  
 plot(as.zoo(all.horizons.training.error[[step]])[,lm.plot[1]], col=plot.colors[1], xlab="", ylab="",main=paste(step\*5,"min forecast"), ylim=c(0.85,1))  
 if(step == 1) mtext("R-squared", side=2, line=0.5, outer=TRUE)  
 # for(i in 2:length(d.plot)) {  
 # lines(as.zoo(all.horizons.training.error[[step]])[,d.plot[i]], col=gray.colors(length(d.plot))[i])  
 # }  
 for(i in 2:length(lm.plot)){  
 lines(as.zoo(all.horizons.training.error[[step]])[,lm.plot[i]], col=plot.colors[i])  
 }  
}  
# Legend  
par(mar=c(0, 0, 0, 0))  
plot(0, type="n", axes=FALSE, xlab="", ylab="", xlim=c(0, 10), ylim=c(0, 10))  
legend("center",  
 legend=c(paste(seq(1:length(grep("LM", colnames(all.horizons.training.error[[step]])))), "day")),  
 col=plot.colors, lty=1, xpd = TRUE, title=substr(file,10,nchar(file)), horiz = TRUE  
 )  
}

## Model testing error

 Figure 6. Testing RMSE and R-squared for each model of each test configuration





##### Page break

Brosilow, Coleman, and Babu Joseph. 2002. “Techniques of Model-Based Control.” In. Prentice Hall Professional.

Burger, Gillian, Cole Sigmon, Christopher Marks, Conor Tyler, Christopher Bye, Daniela Conidi, and Kevan Brian. 2019. “Modeling Aeration Performance for Energy Reduction.” In *Proceedings of the Water Environment Federation*. Vol. 2019.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.

Visioli, Antonio. 2006. *Practical Pid Control*. Springer Science & Business Media.