LLNL-TR-662018

Comparison of multi-physics and multi-analysis ensemble prediction skill for wind ramp events

LDRD Project #12-ERD-069 - FY14 Technical Report

Matthew Simpson
Vera Bulaevskaya
Sonia Wharton
Wayne Miller

Lawrence Livermore National Laboratory
Livermore, CA 94550

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Executive Summary

Results from a comparison of different ensemble forecast methodologies for short-range prediction of wind ramp events using the Weather Research and Forecast (WRF) model are presented. We perform wind ramp forecasts using both a multi-analysis and a multi-physics ensemble philosophy at two study sites with different terrain complexity. The multi-analysis and the multi-physics ensemble forecast strategies are shown to have a high rate of wind ramp detection at both the flat and complex terrain sites. Results indicate the multi-analysis ensemble approach is able to forecast wind ramps more accurately at the flat terrain study site while the multi-physics performs slightly better at the complex terrain site. Both ensemble forecast methods tend to systematically predict the onset of wind ramp events too early at both study sites. However, the ensemble methods perform better at timing the end of a ramp event. The two ensemble types have similar skill in terms of predicting the magnitude of change in power generation associated with the ramp events. At the flat terrain site, the power generation magnitude error tended to be smaller than at the complex terrain site using either ensemble approach. While there was virtually no bias in generation magnitude prediction at the flat site, there was a slight tendency to over-predict the magnitude (by up to 7% of the capacity) at the complex site. An analysis of the model forecast results indicates both the multi-analysis and the multi-physics ensemble methodologies are reasonable approaches for wind ramp events in the context of short-range forecasting.

1. Introduction

Wind ramp events are characterized by a significant change in wind resources over a small time window. Though numerous interpretations of a wind ramp exist, a commonly used definition is a minimum of 50% change in wind park power generation as a fraction of capacity within a period of 4 hours. Wind ramp events create problems for utility grid operators due to the rapid change in renewable generation resources available to the grid. During an up-wind ramp event, renewable generation available to the grid rapidly increases and may need to be curtailed based on utility grid capacity constraints. Conversely, during a down-wind ramp event, renewable generation available to the utility grid rapidly decreases. The sudden disappearance of generation during a down-ramp event creates problems for grid operators since they must find new sources of generation to replace the loss due to the wind ramp. The grid operator's options include adding power from 'spinning reserves' or importing power, potentially at great cost.

Accurate forecasts of wind ramp events benefit grid operators since it provides sufficient time to schedule spinning reserve resources to be made available or to purchase electricity in advance, at lower rates than if done after a down-wind ramp has already started. Unfortunately, forecasting wind ramps is extremely difficult due to the chaotic, intermittent nature of the wind. In addition, many wind parks are located in complex terrain, which makes wind forecasts even more prone to large errors. An ensemble methodology has been proposed to improve wind ramp predictions by taking a probabilistic approach to forecasting. The motivation for the ensemble approach is to account for model uncertainty and quantify the possible future states of the atmosphere in a probabilistic manner (e.g. through the probability density function, PDF) (Mullen, 1994). Two of the major sources of uncertainty that are addressed by different ensemble modeling approaches are errors in the initial / lateral boundary conditions due to sensor error and incomplete observation coverage and errors in the model physics.

A multi-physics (MP) ensemble involves running an atmospheric model over the same period but using different model physics configurations. The motivation for using an MP ensemble is that no single model physics configuration will always have superior performance over a specific geographic region for all plausible atmospheric conditions, making a probabilistic approach necessary to capture the state of the atmosphere as accurately as possible. An MP ensemble approach has been demonstrated as a viable method to account for model parameterization uncertainty and to generate a probabilistic atmospheric forecast (Berner, 2011; Hacker, 2011). MP ensembles are particularly useful for short-range forecasting since the contribution of model physics error to overall forecast uncertainty grows at a rate of two to six times faster than the initial condition uncertainty in the first 12 hours of a weather forecast (Strensrud 2000). Modern prognostic weather forecast models typically have numerous physics choices available for each parameterization category (e.g. land surface model, planetary boundary layer mixing), thereby making running an MP ensemble plausible.

A multi-analysis (MA) ensemble approach, on the other hand, consists of running multiple independent weather forecast simulations over the same period but using different analysis fields (i.e. initial conditions). The goal of the MA ensemble is to account for uncertainty in the initial conditions by sampling across plausible states of the atmosphere. Uncertainty in analysis fields for weather modeling exists for two primary reasons. The first is due to observation error (or sensor error). Surface-based meteorological instruments typically have low sensor error while remotely sensed data, such as from a satellite, will have large observation errors that cannot be ignored. The second primary source of uncertainty in analysis fields is incomplete data coverage, which requires making a best estimate of the state of the atmosphere (referred to as the "background state") in data-sparse regions. By addressing uncertainty associated with observation

error and the background state, a suite of plausible initial conditions can be obtained to provide input to an MA ensemble forecast system.

Project Goal

The goal of this project is to determine whether a multi-analysis or a multi-physics ensemble approach is superior in the short-range (< 30 hours) prediction of wind ramp events. In addition, we will investigate the sensitivity of the ensemble forecast approaches to terrain complexity, by considering two sites with different terrain characteristics. Determining the optimal ensemble forecast strategy for wind ramp prediction is highly beneficial since running both ensemble methodologies is computationally expensive and difficult to operationally maintain. Knowing which forecast strategy is optimal in terms of ramp prediction accuracy is highly valuable as it can lead to large cost savings for wind park and utility grid operators.

Methodology

The methodology used to determine the most accurate ensemble forecast approach for wind ramp prediction consisted of six primary steps. The first step was to identify study locations that were ideal for achieving the project goals. A description of the study locations chosen for this study and the reasons for their selection is provided in Section 2. The second step was to convert observed wind speeds at the study locations into turbine power generation in order to identify power generation ramp case study periods. Meteorological data used in this study and the calculation of turbine generation at the study locations is discussed in Section 3. The third step was running multi-analysis and multi-physics ensemble forecast simulations using a high-resolution prognostic weather model for numerous wind ramp events at the case study locations. The atmospheric model used for this study and the required data inputs are discussed in Section 4. The fourth step was to convert the model predictions of wind speed to power, similarly to what was done in step 2 for the observed wind speeds. The fifth step involved identifying ramps in power trajectories obtained from both observed and ensemble-predicted wind speeds. This is discussed in Section 5. The final step in the methodology was to perform a statistical comparison of the multi-analysis and multi-physics ensemble forecast results at the study locations. Results of the wind ramp ensemble forecast analysis are discussed in Section 6.

2. Study Locations

One of the goals of this project was to compare multi-analysis and multi-physics ensemble forecast results in varying terrain. Therefore, potential study sites were initially screened based on terrain complexity (e.g. flat terrain versus steep ridgelines). In addition, potential study locations were only considered if the meteorological data necessary to calculate turbine generation was available. The final criterion for the site selection was the presence of renewable resources in the surrounding area so that the study results would be relevant.

The first selected study location was the Chisholm View Wind Park, which is located in north-central Oklahoma. Chisholm View is located in a region of flat terrain with elevation slopes predominately below 3°. The land use in the Chisholm View region is mostly cropland and grassland making it representative of large portions of the central United States. Numerous wind turbines are in the region making numerical modeling results relevant to an existing renewable generation community.

The second study domain selected is the Lawrence Livermore National Laboratory (LLNL) Site 300 complex, which is a high-explosive test facility maintained by the Department of Energy (DOE). Site 300

is located in north-central California near the Altamont Pass portion of the Diablo Range. Site 300 is characterized by steep ridgelines resulting in highly complex terrain. Site 300's proximity to the Altamont Pass, which is home to one of the largest wind parks in California, will make study findings relevant to a large portion of existing renewable generation resources.

3. Data

Atmospheric observations from meteorological towers at the two study locations were used to calculate turbine power generation and to verify the atmospheric model prediction results. In addition, gridded meteorological data were used to initialize and provide lateral boundary conditions for the wind ramp forecast simulations. Publically available observations from a variety of meteorological networks were also used for initialization of the weather forecasts. A description of the different meteorological data used in this study is provided below.

3.1 Chisholm View tower data

Two meteorological towers are located at the Chisholm View wind park and were both used for this study. Due to the proximity of the Chisholm View meteorological towers to a wind turbine array, only specific wind sectors were considered wake-free at each observation tower. We used the wind speed observations from the 'north' meteorological tower to calculate turbine power generation when the observed wind direction at the tower was between 270° and 80°. When the observed wind direction was from 80° to 270°, we used wake-free observations from the 'south' meteorological tower. Wind speed observations made at a height of 80 m above the surface were used to calculate turbine generation using GE 1.68 MW power curve data since these are common turbine specifications in the region. Chisholm View weather observations at both meteorological towers were available at 10-minute intervals over the period of November 2013 through January 2014.

3.2 LLNL – Site 300 tower data

A single meteorological tower is located in the north-central region of the LLNL – Site 300 complex with wind speed / direction observation heights at 10, 23, and 52 m above ground level. Since no wind turbines are located at Site 300, a wake-free wind sector filter was not needed for data processing. Wind speed data at 52 m height and the power curve for a Mitsubishi MWT – 1000 kW turbine were used to estimate turbine generation. Atmospheric data over the period of October 2013 until July 2014 were downloaded from the Site 300 data public website at 15-minute intervals.

3.3 Gridded atmospheric data

Gridded analysis fields provided by the Global Ensemble Forecast System (GEFS) atmospheric model are used to initialize and update boundary conditions for the weather forecasts. GEFS data are available at 6-hour intervals at one-degree (~ 110 km) horizontal resolution. GEFS data were downloaded from the NOAA Operational Model Archive and Distribution System (NOMADS) data portal (http://nomads.ncep.noaa.gov/pub/data/nccf/com/gens/prod/). Gridded atmospheric data are available at 27 pressure levels in the GEFS data set. GEFS atmospheric variables used to perform the prognostic weather simulations include atmospheric pressure, geopotential height, horizontal wind components, air temperature, specific humidity, sea level pressure, and sub-surface temperature and moisture.

The GEFS modeling system consists of a total of 21 unique ensemble members. One of the ensemble members is considered to be a 'control' run by using an unperturbed, best estimate of the current state of the atmosphere. A Breeding Method (BM) is utilized to generate perturbed initial conditions for the remaining

20 ensemble members to account for uncertainties in the original weather observations and background state.

3.4 MADIS data

Weather observations from the Meteorological Assimilation Data Ingest System (MADIS; Miller 2005, Miller 2009) were utilized during a 6-hour dynamical initialization period for the weather forecasts (details of the forecast initialization are provided in the next section). Specifically, weather observations from the metar, maritime, mesonet, profiler, and raob data networks were used for the forecast model initialization. Atmospheric variables from MADIS that were assimilated by the forecast model include atmospheric pressure, geopotential height, horizontal wind vectors, air temperature, and relative humidity.

4. Atmospheric model

The non-hydrostatic, fully compressible Weather Research and Forecast (WRF) (Skamarock et al., 2008) numerical weather prediction (NWP) model was used to generate wind and power time series forecasts for this project. Specifically, we used version 3.5.1 of the advanced research dynamical core of the WRF code. WRF is a state-of-the-science community-based atmospheric model deigned to address both operational weather forecasting and broad atmospheric research needs. WRF was developed via collaboration among numerous academic, research, and government organizations to streamline the transfer of atmospheric research findings to an operational capacity. The large set of available model physics schemes and data assimilation approaches coupled with efficient model nesting make WRF appropriate for performing simulations on scales of motion from tens of meters to thousands of kilometers. Source code for WRF is available for download by the public at no cost through the National Center for Atmospheric Research (NCAR) supported website[1].

Several features make WRF the ideal atmospheric model to use for investigating the skill of different ensemble forecasting approaches for wind ramp events. WRF is currently widely used by both research and private weather forecasting sectors. Therefore, any 'best practices' concluded from this study can be utilized by an existing large user base. In addition, WRF has numerous schemes available for each of the main physics categories, leading to a large number of working physics configurations, which is ideal for any ensemble-based study. WRF can be initialized with different initial and lateral boundary conditions for the same period, making a multi-analysis ensemble forecast approach possible. Also, WRF source code has been developed and optimized to run on high-performance parallel computing resources, allowing for a large number of high-resolution ensemble forecasts to be performed in a reasonable timeframe.

Model Domains

A WRF triple-nested model domain configuration was used for the atmospheric forecasts at both Chisholm View and Site 300. The geographic coverage of the WRF domains centered on the Chisholm View wind park is shown in Figure 1a while the Site 300 WRF domain configuration is shown in Figure 1b. The same domain configuration (e.g. grid spacing, grid points) was used for both study locations. The outermost model domain, which is labeled D1, has a horizontal grid spacing of 27 km and covers a large region surrounding the study locations. A large outer model domain was necessary given the coarse resolution of the GEFS input data. Nested model domains 2 and 3 have horizontal grid spacing of 9 and 3 km, respectively. The innermost model domain (labeled D4) has a horizontal grid spacing of 1 km, which is

[1] http://www2.mmm.ucar.edu/wrf/users/download/get_source.html

sufficiently high resolution to model small-scale atmospheric features present at both study locations. It should also be noted that no cumulus parameterization was utilized in domains 3 and 4 due to their high resolution.

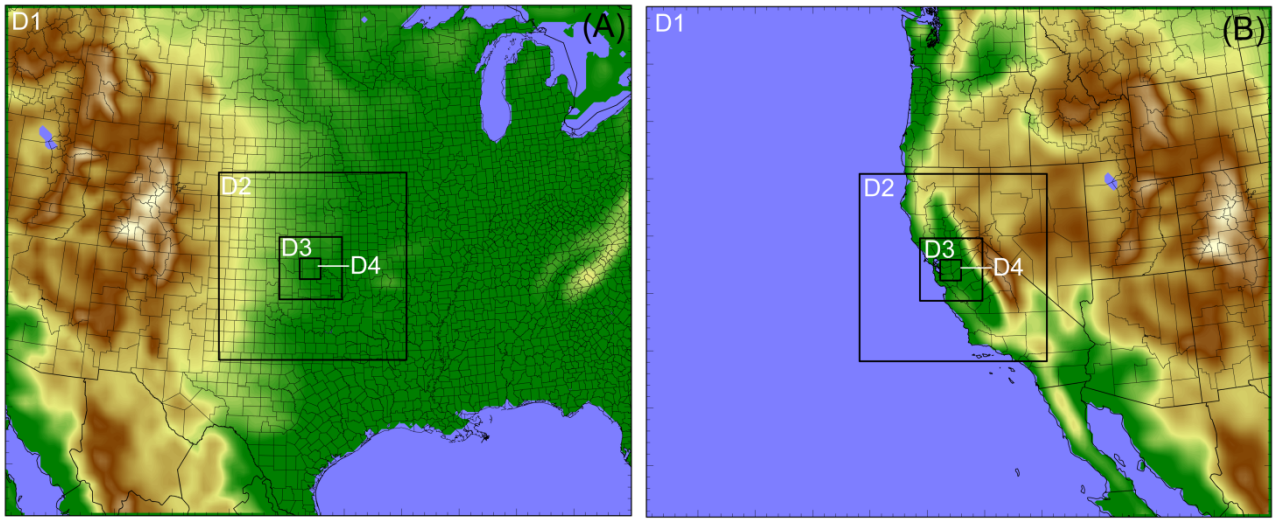

Figure 1. (a) Nested WRF domain configuration used for the atmospheric ensemble modeling study centered on the Chisholm View wind park. Model domain 1 (labeled D1) and domain 2 (D2) have a horizontal grid spacing of 27 and 9 km, respectively. Model domains 3 and 4 (D3, D4) have a horizontal grid spacing of 3 and 1 km, respectively. (b) Same as (a) but centered on the LLNL Site 300 complex-terrain study location.

We used a total of 50 terrain-following vertical sigma levels for the WRF ensemble forecasts. The vertical resolution in the lowest 200 m of the atmosphere for the WRF grid was roughly 15-20 m. This degree of high vertical resolution was necessary since the focus of the forecasts was hub-height wind speeds, from which the turbine generation could be calculated. The vertical resolution of the WRF grid was gradually increased above 200 m until the top of the model grid at 50 hPa (roughly 20 km above sea level). Due to the high horizontal and vertical resolution of the WRF grid, a small fixed numerical time step of 5 seconds for the innermost domain was used for the ensemble forecast runs. WRF model output was saved every 10 minutes for the Chisholm View forecasts to be consistent with the observation data frequency and every 15 minutes for the Site 300 ensemble runs.

Ensemble design

The multi-analysis forecast runs consisted of a total of 21 ensemble members. Each ensemble member used a unique set of initial and lateral boundary conditions from the GEFS input data described in Section 3.3. Since typically only the model initial condition fields are varied in a multi-analysis ensemble, we used the following single set of model physics options for all of the multi-analysis ensemble runs: MYJ planetary boundary layer model, NOAH land surface model, Kain-Fritsch cumulus parameterization, Dudhia shortwave radiation physics, RRTM longwave radiation physics, and WSM 5-class moist physics. This particular set of physics options was chosen since it represents a configuration commonly used by the atmospheric modeling community.

The multi-physics forecast runs also consisted of 21 ensemble members. Each ensemble member used a unique configuration of model physics options. A list of the physics options utilized by each ensemble

member for the WRF forecasts is provided in Table 1. The physics configurations were developed by varying options in all of the primary physics categories. In particular, we varied the land surface model and planetary boundary layer the most since previous research has shown these two physics categories to have the greatest impact on simulated near-surface wind fields (Lee et al. 2012, Simpson et al. 2014). GEFS 'control' analysis fields and lateral boundary conditions were used to initialize all of the multi-physics ensemble member forecast runs.

Table 1: Configurations of WRF Multi-physics Ensemble Members

Member	PBL	LSM	Cumulus	Microphysics	Longwave	Shortwave
1	YSU	Thermal	BMJ	Ferrier	RRTM	Dudhia
2	YSU	Noah	Grell	WSM6	RRTM	Dudhia
3	YSU	RUC	KF	Thompson	RRTM	Goddard
4	YSU	PX	KF	WSM5	CAM	Dudhia
5	MYJ	Thermal	KF	WSM5	RRTM	Goddard
6	MYJ	Noah	BMJ	Ferrier	RRTM	Dudhia
7	MYJ	Noah	KF	WSM6	RRTM	CAM
8	MYJ	RUC	KF	Lin	CAM	CAM
9	QNSE	Noah	KF	WSM6	RRTM	RRTMG
10	QNSE	PX	Grell	Ferrier	CAM	Dudhia
11	QNSE	RUC	GD	Ferrier	CAM	Dudhia
12	MYNN	Thermal	KF	Lin	RRTM	Goddard
13	MYNN	RUC	Grell	WSM6	RRTM	Dudhia
14	MYNN	Noah	BMJ	Ferrier	RRTM	RRTMG
15	ACM2	PX	BMJ	WSM5	RRTM	CAM
16	BouLac	RUC	Grell	Lin	RRTM	Dudhia
17	BouLac	Noah	BMJ	Thompson	CAM	RRTMG
18	BouLac	PX	GD	Thompson	CAM	RRTMG
19	UW	Noah	Grell	Ferrier	CAM	Goddard
20	UW	Thermal	BMJ	Lin	RRTM	RRTMG
21	UW	RUC	KF	Lin	RRTM	Dudhia

Four-dimensional data assimilation

A dynamical initialization 6-hour spin-up period based on the principles of four-dimensional data assimilation (FDDA) was used on all of the WRF ensemble forecasts. Both the WRF analysis (Stauffer and Seaman 1994) and observational (Liu et al. 2005, Liu et al. 2009) FDDA nudging were applied during the pre-forecast spin up period. The goal of the FDDA spin up period was to nudge the numerical integration towards an observed state via relaxation terms. By doing so, the model is as close as possible to the 'true' state of the atmosphere when the pure forecast begins, thereby leading to more accurate predictions. The FDDA analysis nudging term acts to constrain large-scale motion while the observational nudging influences the simulation of small-scale atmospheric features. Observational FDDA nudging was utilized during the spin-up period in WRF model domains 2, 3, and 4 while the analysis nudging was used only above the boundary layer in model domains 1 and 2.

Case studies

Using the meteorological observations at Chisholm View and Site 300 and the corresponding power curve previously described in Sections 3.1 and 3.2, turbine generation data were calculated over the entire period of wind speed data availability. We selected 8 ramp event case studies for each site using the calculated power data and the definition of a wind ramp as a minimum of 50% change in turbine generation as fraction of capacity over a period of 4 hours or less. Forecast start and end times for the Chisholm View and Site 300 wind ramp case studies are shown in Tables 2 and 3, respectively. Note that an "Up" or a "Down" entry in the Ramp type column of the two tables implies the presence of one ramp, while an "Up-down" entry implies the presence of 2 ramps (there were no cases with a down-up ramp). All of the forecast simulations were a total of 30 hours in duration after excluded the 6-hour spin-up period.

Table 2: Chisholm View wind park case study periods.

Case study	Forecast start	Forecast end	Ramp type
1	2013-11-12, 18Z	2013-11-14, 00Z	Up
2	2013-11-21, 00Z	2013-11-22, 06Z	Up
3	2013-11-28, 00Z	2013-11-29, 06Z	Down
4	2013-11-28, 12Z	2013-11-29, 18Z	Up-down
5	2013-12-03, 18Z	2013-12-05, 00Z	Up
6	2013-12-10, 12Z	2013-12-11, 18Z	Up-down
7	2013-12-24, 12Z	2013-12-25, 18Z	Down
8	2014-01-11, 12Z	2014-01-12, 18Z	Up

Table 3: Site 300 wind park case study periods.

Case study	Forecast start	Forecast end	Ramp type
1	2014-01-30, 12Z	2014-01-31, 18Z	Down
2	2014-02-06, 00Z	2014-02-07, 06Z	Up-down
3	2014-03-01, 12Z	2014-03-02, 18Z	Down
4	2014-04-03, 12Z	2014-04-04, 18Z	Up-down
5	2014-04-19, 12Z	2014-04-20, 18Z	Down
6	2014-05-06, 18Z	2014-05-08, 00Z	Down
7	2014-05-29, 18Z	2014-05-31, 00Z	Down
8	2014-06-02, 12Z	2014-06-03, 18Z	Up-down

Ensembles of power

The WRF ensembles of wind speed predictions described above were converted to power forecasts using the power curves mentioned in Sections 3.1 and 3.2, which were used to convert the observed wind speeds to power at the two sites. The resulting power ensembles are shown in Figures 1 and 2 of Appendix A, for each case study and location. As can be seen from the figures, while at Chisholm View the MP ensembles tend to have a larger spread than the MA ensembles for all case studies, at Site 300 this is only true approximately 50% of the time. As might be expected, the larger spread of the MP ensembles at Chisholm

View is associated with greater accuracy: as can be seen in Table 1 in Appendix B, these capture a higher percentage of observations for all case studies than the MA ensembles.

It should be noted that at Chisholm View both ensembles perform poorly in terms of capturing the "true" observations, as reflected in low percent capture values in Appendix B - Table 1. The accuracy of the ensembles is much higher at Site 300, as seen from higher values in Table 2 of Appendix B. Here, too, greater spread tends to be associated with greater accuracy, or percent capture of "true" observations. An important caveat to this discussion is that all the power values in this work are *estimates* based on a power curve, and percent capture results should therefore be interpreted with some degree of caution. Nevertheless, these provide an indication of the two ensemble methods' performance across the two locations.

5. Ramp detection algorithm

As stated earlier, our operational definition of a ramp is a minimum of 50% change in power generation as a fraction of capacity in a period of 4 hours or less. Applying this rule to the data is non-trivial, however, due to the noisy nature of power generation. In particular, determining the onset and end times of a ramp is challenging because even if the trend in power is generally increasing or decreasing over a period of time, the trend may not be purely monotonic. As a result, to mitigate the impact of noise, we first smooth the data using a natural spline (Hastie 1992) and then determine the change-points (time points at which a change in the trend, up or down, occurs) using the smoothed version of power time series. We then use the change points alone to smooth the data again and determine the new change points. We repeat these steps until the number of change points is no longer sufficient to perform smoothing in a reliable manner.

Once the final change points are determined, we then determine whether each segment between two change points satisfies the power ramp definition criterion. If it does, in case of an up ramp, we use time points at which the minimum and the maximum power value in the segment occurred as the onset and the end times of the ramp, respectively. In case of a down ramp, the time points associated with the maximum and the minimum power values are declared onset and end times, respectively. The algorithm above was implemented using R statistical software package (R Core Team, 2014).

While the algorithm allowed automated ramp detection, it was not perfect at correctly determining the onset and end times of each ramp. Thus, some amount of manual adjustment to the onset and end times identified by the algorithm was applied to fine-tune these values. Once the onset and end times were finalized, the difference in power between the onset and end times was taken to be the ramp magnitude (note that the magnitude includes the direction of the ramp).

6. Results

Once the ramps were identified for the power time series based on both the observed wind speed data and the WRF ensemble predictions, we calculated the following metrics for each ensemble method, case study and location:

- Average percent detection of ramps, i.e., the percentage of the total number of true ramps (1 or 2 depending on the case study) detected by each member within the ensemble method, averaged across the ensemble members
- Root mean squared error (RMSE) of the following quantities for the *detected* ramps:

- Onset time predicted by the ensemble member relative to the true onset time for the ramps that are detected
- End time predicted by the ensemble member relative to the true onset time for the ramps that are detected
- Magnitude predicted by the ensemble member relative to the true onset time for the ramps that are detected

These metrics are listed for each location, both by case study and averaged across the case studies, in Tables 4 and 5. Figures 1□4 of Appendix C show examples of successful and poor ramp prediction for some of the ensemble members.

Verification statistics for the Chisholm View multi-physics and multi-analysis ensemble forecasts are shown in Table 4. Both the multi-physics (MP) and multi-analysis (MA) ensembles have a high rate of wind ramp detection. On average, 88% of the MA and 75% of the MP ensemble members accurately predicted the occurrence of wind ramp events at Chisholm View. On an individual case basis, the wind ramp detection rate of the MA and MP methods were comparable. However, case study 1 was an exception to this observation. An analysis of case 1 revealed that the MP ensemble members that utilized the NOAH land surface model (LSM) accurately predicted the wind ramp. Since the MA ensemble utilizes the NOAH LSM, this likely contributed to the 100% wind ramp detection rate for case study 1. Clearly, the physics associated with the NOAH LSM was critical to accurately forecasting the case study 1 wind ramp and highlights the importance of selecting a physics configuration for the MA ensemble that contains physics that will likely perform best in a specific region.

On average, the MA and MP ensembles have a comparable RMSE for the onset timing of the wind ramps of 2.4 and 2.6 hours, respectively. A large amount of variation in the RMSE values is seen among the different case studies. For example, both MA and MP RMSE values for the start of the wind ramp for case 1 are greater than 4 hours while the RMSE values for case study 7 are closer to 1 hour. The Bias for both MA and MP ensembles predicting the timing of ramp onset is -2.0 hours. This implies both ensemble approaches systematically predict the onset of ramps too soon at the Chisholm View wind park. MA and MP RMSE values for the timing of the end of the ramp events are slightly better with values of 2.0 and 2.5 hours, respectively. Like the onset timing RMSE values, a great deal of variation exists among the case studies for the accuracy of predicting the timing of the end of the wind ramps. The average MA and MP Bias values for the timing of the end of wind ramps are both close to zero, which implies there is no systematic error.

In addition to the timing of a wind ramp, the change in turbine generation power associated with the ramp event is another important parameter for a forecast model. The average percent change in generation power RMSE values for the MA and MP ensembles at Chisholm view is 15% and 17%, respectively. The RMSE generation values are comparable but the MA ensemble is shown to perform better in 6 of the 8 case studies indicating some skill. The average Bias in the generation change error is close to 0% for both of the ensemble methodologies. This finding indicates there in no systematic error in over predicting or under predicting the change in power generation during a ramp event at the Chisholm View wind park.

Table 4. Performance metrics for ramp detection and characterization for each case study and ensemble method at Chisholm View.

Case Study	Ensemble Method	Wind Ramp Detection (%)	Ramp Onset RMSE (hr)	Ramp End RMSE (hr)	Δ Power RMSE (%)	Ramp Onset Bias (hr)	Ramp End Bias (hr)	Δ Power Bias (%)
1	Analysis	100	4.2	0.6	14	-4.2	0.5	-14
1	Physics	33	4.5	3.4	18	-2.7	1.1	-14
2	Analysis	100	1.5	2.5	5	-0.7	2.3	1
2	Physics	100	1.4	4.5	4	-0.6	4.5	2
3	Analysis	95	3.2	1.3	15	-3.0	-1.3	14
3	Physics	67	3.3	1.5	16	-2.9	-1.4	11
4	Analysis	45	2.3	2.0	24	-1.7	0.5	-8
4	Physics	41	3.7	1.8	19	-3.3	-0.3	-9
5	Analysis	100	2.7	1.4	12	-2.7	-1.1	-9
5	Physics	95	1.9	1.3	13	-1.8	0.1	-10
6	Analysis	67	2.8	3.3	28	-1.5	-1.6	-2
6	Physics	71	2.9	3.0	30	-1.5	-2.2	-4
7	Analysis	100	0.9	3.0	17	-0.9	-2.9	16
7	Physics	95	1.4	3.0	24	-1.3	-2.9	21
8	Analysis	100	1.8	1.8	5	-1.6	1.6	-1
8	Physics	95	1.8	1.4	11	-1.7	-0.2	-5
Average	Analysis	88	2.4	2.0	15	-2.0	-0.3	0
Average	Physics	75	2.6	2.5	17	-2.0	-0.1	-1

Verification statistics for the Site 300 MA and MP ensemble forecast results are presented in Table 5. Consistent with results from Chisholm View, both the MA and MP ensembles members detect a high percentage of the wind ramp events. On average, 84% of the MA and 88% of the MP ensemble members accurately predicted wind ramp events at Site 300. The wind ramp detection rate of the MA and MP ensemble members were reasonably comparable across the case studies except for case 4. The case 4 wind ramp was predicted by 86% of the MP ensemble members while only 41% of the MA members correctly predicted the ramp. The exact cause of why the MA ensemble performed so poorly for case 4 is not obvious and will be the focus of additional analysis in the future.

Contrary to results at Chisholm View, the MP ensemble more accurately predicts the timing of ramp event onset at Site 300. The MA and MP ensemble RMSE values for the wind ramp start timing is on average 4.2 and 3.7 hours, respectively. A large degree of variation exists among the case studies regarding the prediction error of ramp onset. For example, the ramp onset timing error for both ensemble methods is around 2 hours for case 5 while errors around 8-9 hours are associated with case 7. The average Bias for MA and MP ensemble runs for predicting the timing of ramp onset is -3.0 and -2.5 hours, respectively. This indicates both ensemble methods at Site 300 are systematically predicting wind ramps will start too early. RMSE values for the timing of the end of the ramp events are better by around an hour with MA and MP ensemble values of 3.0 and 2.8 hours, respectively. The average MA and MP Bias values for the timing of the end of wind ramps exhibit less systematic Bias than prediction of the ramp onset timing. MA and MP Bias values for wind ramp ending timing are both -1.4 hours.

The average percent change in generation power RMSE values at Site 300 is slightly higher than at Chisholm View wind park. MA and MP ensemble RMSE error values for the change in power generation is 20% and 19%, respectively, indicating both methods do a reasonably good job at predicting the magnitude change in generation during a wind ramp at Site 300. The lowest change in generation RMSE value is 4%, which is from the MP ensemble for case 2. However, the largest power change RMSE value is 33%, which is also associated with the MP ensemble but valid for case 8. The average Bias in the generation change error is around 5-7% for both of the ensemble methodologies. This indicates there is a small systematic tendency to overestimate the change in power generation during a wind ramp event at Site 300.

Table 5. Performance metrics for ramp detection and characterization for each case study and ensemble method at Site 300.

Case Study	Ensemble Method	Wind Ramp Detection (%)	Ramp Onset RMSE (hr)	Ramp End RMSE (hr)	Δ Power RMSE (%)	Ramp Onset Bias (hr)	Ramp End Bias (hr)	Δ Power Bias (%)
1	Analysis	86	5.0	5.0	21	-4.2	-3.9	15
1	Physics	100	4.4	4.5	16	-4.0	-4.1	14
2	Analysis	86	5.7	5.2	19	-4.8	-4.4	-2
2	Physics	100	5.3	4.2	4	-5.1	-4.0	-1
3	Analysis	100	3.7	1.9	13	-3.4	0.3	7
3	Physics	100	3.4	2.2	18	-3.1	-0.4	13
4	Analysis	41	1.9	1.6	29	1.3	0.4	-6
4	Physics	86	1.6	0.8	24	1.3	0.2	-3
5	Analysis	95	2.1	1.5	16	-2.0	-1.2	13
5	Physics	100	1.9	1.5	17	-1.8	-1.2	15
6	Analysis	100	4.7	1.9	30	-3.6	-0.5	24
6	Physics	85	2.9	2.3	23	-1.8	0.5	19
7	Analysis	100	9.0	4.1	16	-9.0	-4.0	-14
7	Physics	95	7.9	4.4	16	-7.7	-4.3	-5
8	Analysis	62	2.0	2.7	18	1.8	1.9	0
8	Physics	36	2.4	2.6	33	2.1	1.7	4
Average	Analysis	84	4.2	3.0	20	-3.0	-1.4	5
Average	Physics	88	3.7	2.8	19	-2.5	-1.4	7

6. Conclusions

Based on our analysis of the WRF ensemble forecast results, we make the following conclusions:

- the choice of model physics configuration is important when designing a multi-analysis forecast system
- the multi-analysis and the multi-physics ensemble methodologies have a high rate of wind ramp detection in both flat and complex terrain
- the multi-analysis ensemble approach is able to forecast wind ramps more accurately at the Chisholm View wind park (flat terrain)
- the multi-physics ensemble approach is able to forecast wind ramps slightly more accurately at Site 300 (complex terrain)
- both the multi-analysis and the multi-physics ensemble tend to predict wind ramp event onset too early
- the two forecast ensemble methods have similar skill in terms of predicting the change in power generation associated with ramp events
- both forecast ensemble methods tend to predict the timing of when a wind ramp will end more accurately than the time of onset
- both the multi-analysis and the multi-physics ensemble methodologies are reasonable approaches for wind ramp event short-range forecasting

7. Future Work

The results presented in this report provide valuable insight to the ability of a multi-physics and multi-analysis ensemble to predict wind ramp events in flat and complex terrain. However, more forecast simulations of wind ramp events are needed to increase the confidence in the statistical analysis of the forecast results. Also, ideally the verification statistics in the future would be based on observed turbine generation data versus data calculated using a power curve function.

In addition, since the onset and ending time of a wind ramp event can be subjective at times, a more sophisticated ramp detection algorithm is likely needed for projects in the future. A greater understanding of the choice of physics configuration for the multi-analysis ensemble is also needed to ensure the statistical relevance of ramp prediction analysis results.

References

Berner, J., S. Y. Ha, J. P. Hacker, A. Fournier, C. Snyder, 2011. Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations, *Monthly Weather Review* 139, pp. 1972–1995.

Hacker, J. P., S. Y. Ha, C. Snyder, J. Berner, F. A. Eckel, E. Kuchera, M. Pocernich, S. Rugg, J. Schramm, and X. Wang, 2011. The U.S. Air Force Weather Agency's Mesoscale Ensemble: Scientific Description and Performance Results. *Tellus 63A*, pp. 625-641.

Hastie, T. J. Generalized additive models. Chapter 7 of *Statistical Models in S* eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole, 1992.

Lee, J.A., W.C. Kolczynski, T.C. McCandless, and S.E. Haupt, 2012. "An Objective Methodology for Configuring and Down-Selecting an NWP Ensemble for Low-Level Widn Prediction." *Monthly Weather Review* 140, 2270-2286.

Liu, Y., A. Bourgeois, T. Warner, S. Swerdlin and J. Hacker, 2005. An Implementation of Obs-nudging-based FDDA into WRF for Supporting ATEC Test Operations, *2005 WRF User Workshop*, Paper 10.7.

Liu, Y., T. Warner, W. Wu, G. Roux, W. Cheng, Y. Liu, F. Chen, L. Delle Monache, W. Mahoney and S. Swerdlin, 2009. A versatile WRF and MM5-based weather analysis and forecasting system for supporting wind energy prediction. 23rd WAF / 19th NWP Conf. , AMS, Omaha, NE. June 1- 5, 2009. Paper 17B.3

Miller, P.A., M. Barth, L. Benjamin, D. Helms, M. Campbell, J. Facundo, and J. O'Sullivan, 2005. The Meteorological Assimilation and Data Ingest System (MADIS): Providing Value-added Observations to the Meteorological Community, *21st Conference on Weather Analysis and Forecasting*, Washington, D.C., American Meteorological Society, 2005.

Miller, P.A., M.F. Barth, L.A. Benjamin, 2009. A 2009 Update on the NOAA Meteorological Assimilation Data Ingest System (MADIS), *25th Conference on International Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Phoeniz, AZ, American Meteorological Society, 2009.

Mullen, S. L., and D. P. Baumhefner, 1994. Monte Carlo Smulations of Explosive Cyclogenesis, *Monthly Weather Review 122*, pp. 1548–1567.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Simpson, M., V. Bulaevskaya, P. Sotorrio, and T. Edmunds, 2014. Determination of the Optimal Multi-physics Ensemble Configuration for Predicting California Renewable Generation. Lawrence Livermore National Laboratory Technical Report, LLNL-TR-657738.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X-Y. Huang, W. Wang and J. G. Powers, 2008. A Description of the Advanced Research WRF Version 3. NCARTech Note, NCAR/TN–475+STR, 113 pp.

Stauffer, D. R., and N.L. Seaman, 1994. On Multi-Scale Four-Dimensional Data Assimilation, *Journal of Applied Meteorology 33*, pp. 416-434.

Stensrud, D., J. W. Bao, and T. T. Warner, 2000. Using Initial Condition and Model Physics Perturbations in Short-range Ensemble Simulations of Mesoscale Convective Systems, *Monthly Weather Review 128*, pp. 2077–2107, 2000.

Appendix A. Plots of power ensembles for each location and case study

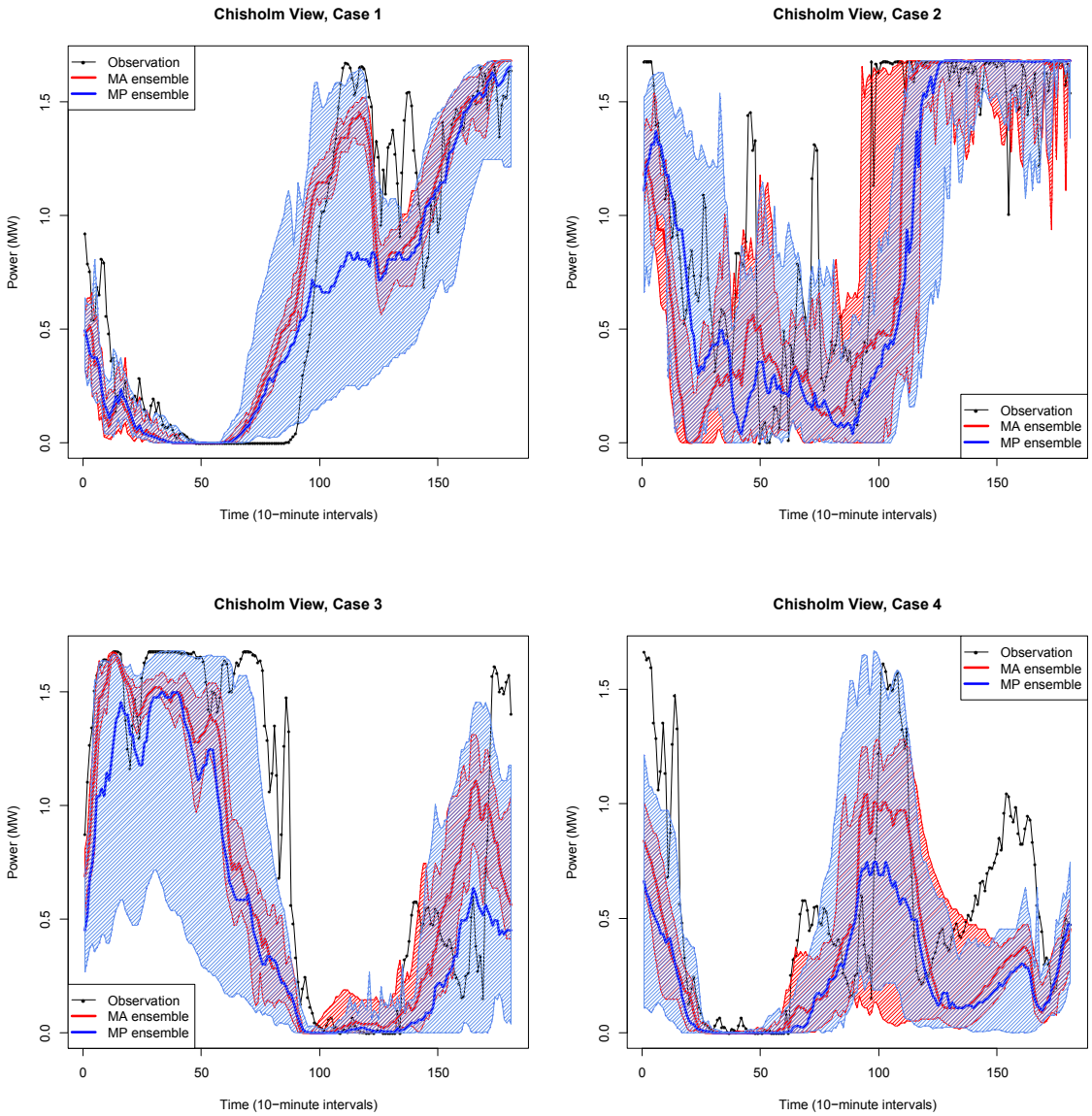

Figure 1(a). Resulting MA (in red) and MP (in blue) ensembles of power predictions for Chisholm View for case studies 1—4. The lower and upper limits of the shaded areas for each ensemble method are the ensemble minimum and the maximum power value, respectively, at each time point. The solid curve in the middle of the shaded area represents the median of the ensemble at each time point. The power values based on observed wind speed data are shown as black circles.

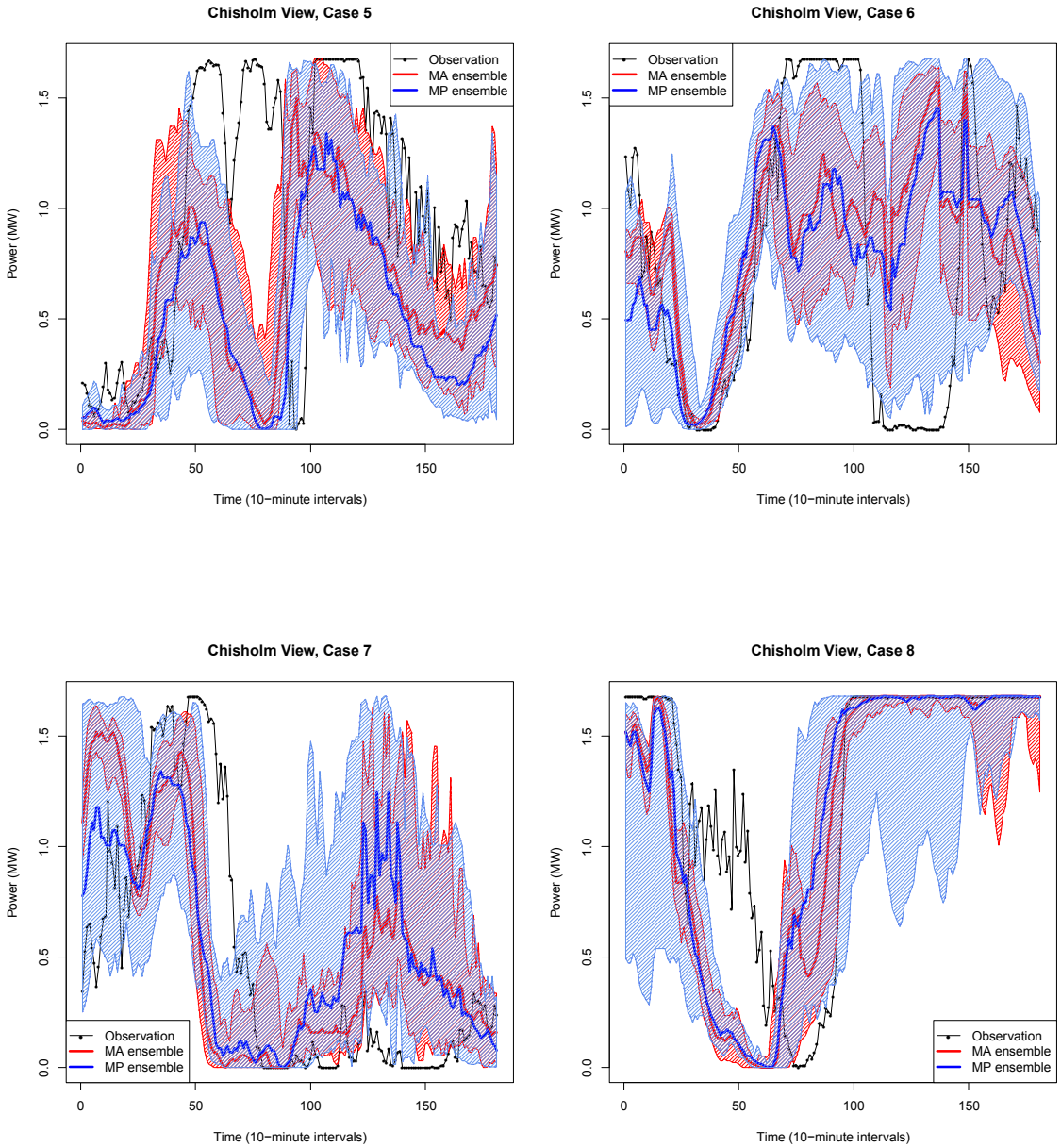

Figure 1(b). Same as Figure 1(a), but for Chisholm View case studies 5—8.

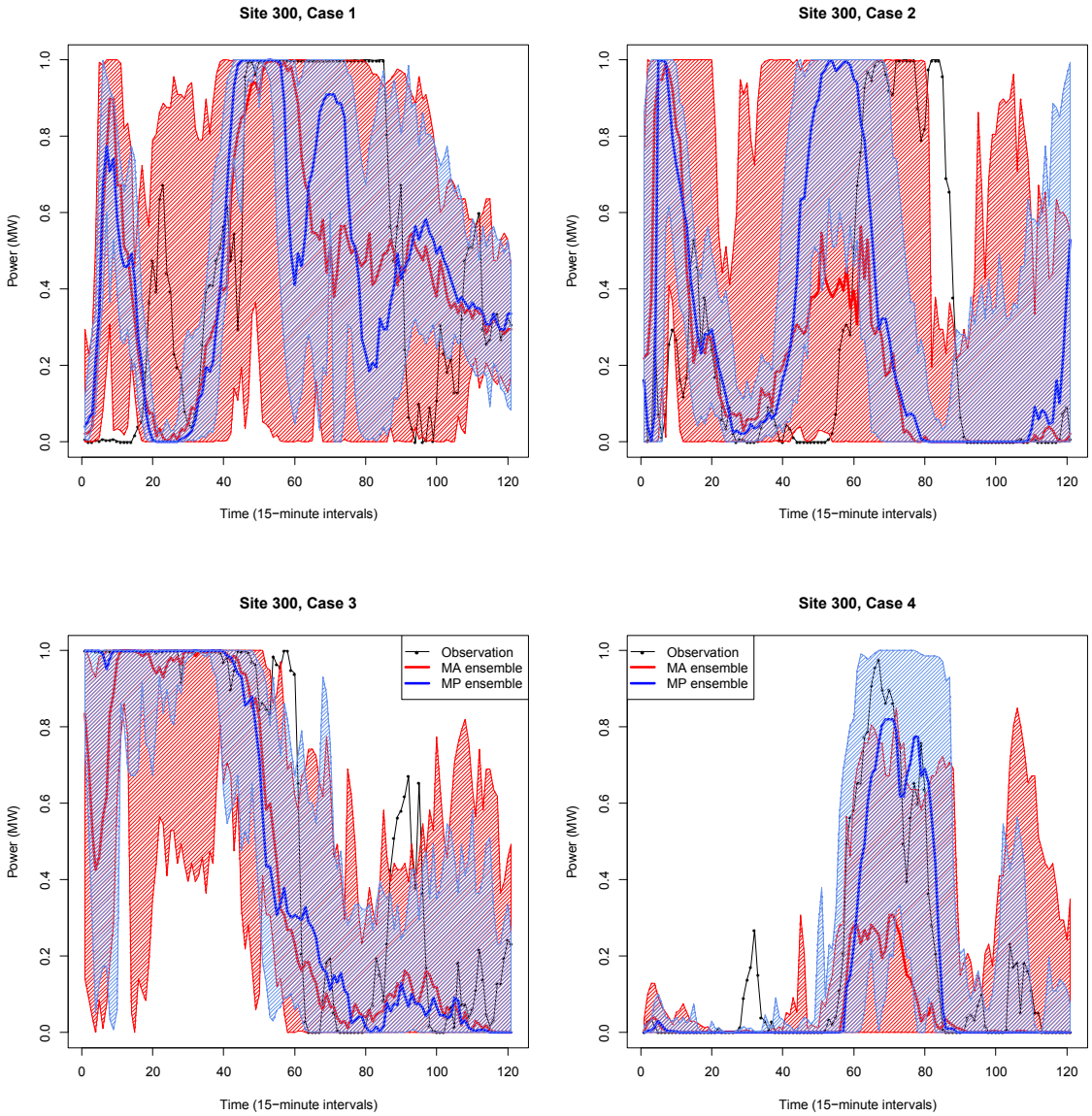

Figure 2(a). Same as Figure 1(a), but for Site 300 case studies 1—4.

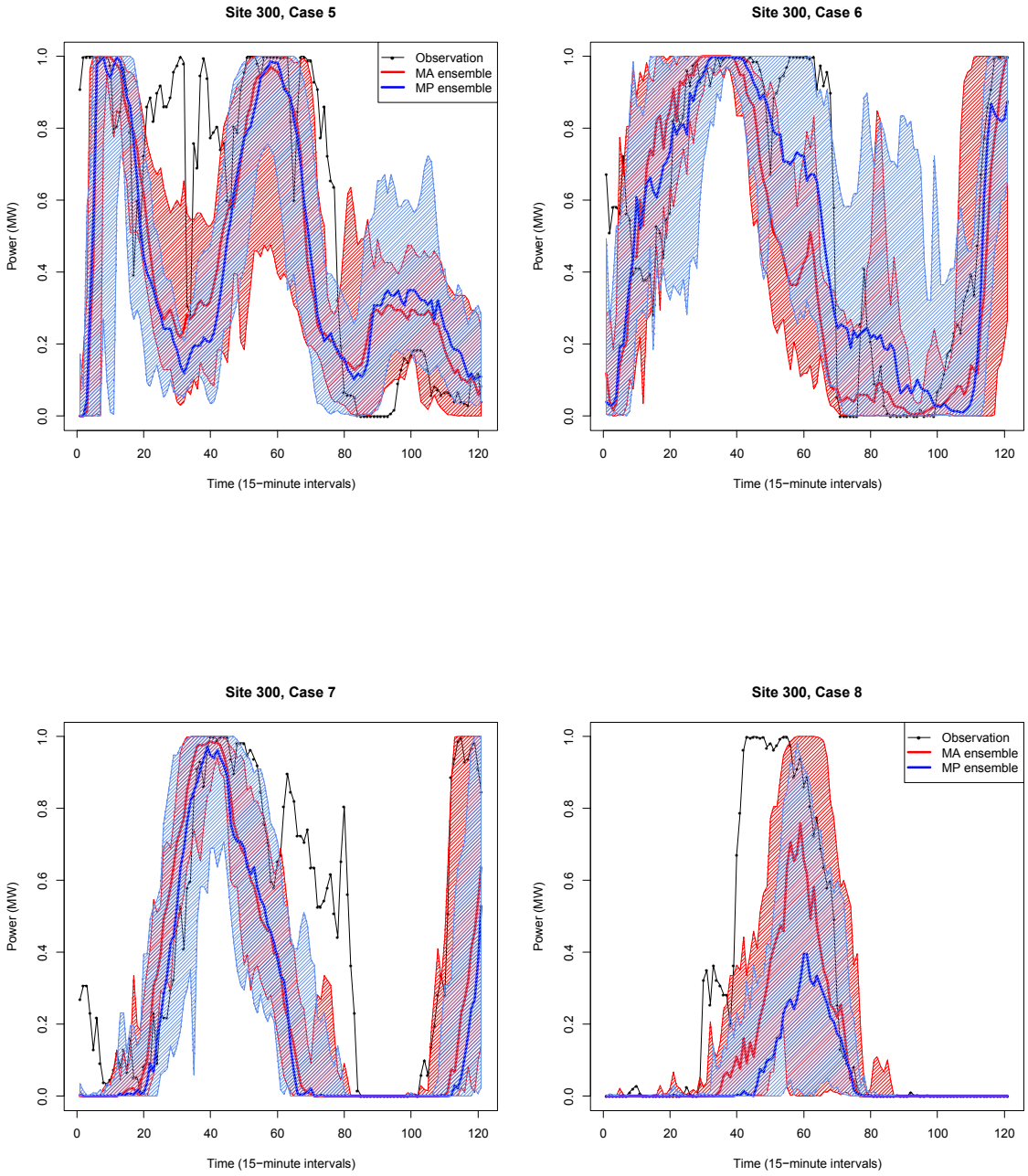

Figure 2(b). Same as Figure 1(b), but for Site 300 case studies 5—8.

Appendix B. Ensemble accuracy

Table 1. Percent of observations in the Chisholm View power series in each case study captured by each ensemble method.

Case	Observation capture (%)	
	MA Ensemble	**MP Ensemble**
1	33	58
2	64	75
3	34	51
4	44	54
5	29	35
6	27	46
7	32	52
8	52	61

Table 2. Same as Table 1, but for Site 300.

Case	Observation capture (%)	
	MA Ensemble	**MP Ensemble**
1	79	52
2	78	60
3	83	83
4	78	89
5	54	46
6	65	83
7	53	55
8	69	63

Appendix C. Examples of ensemble member ramp prediction

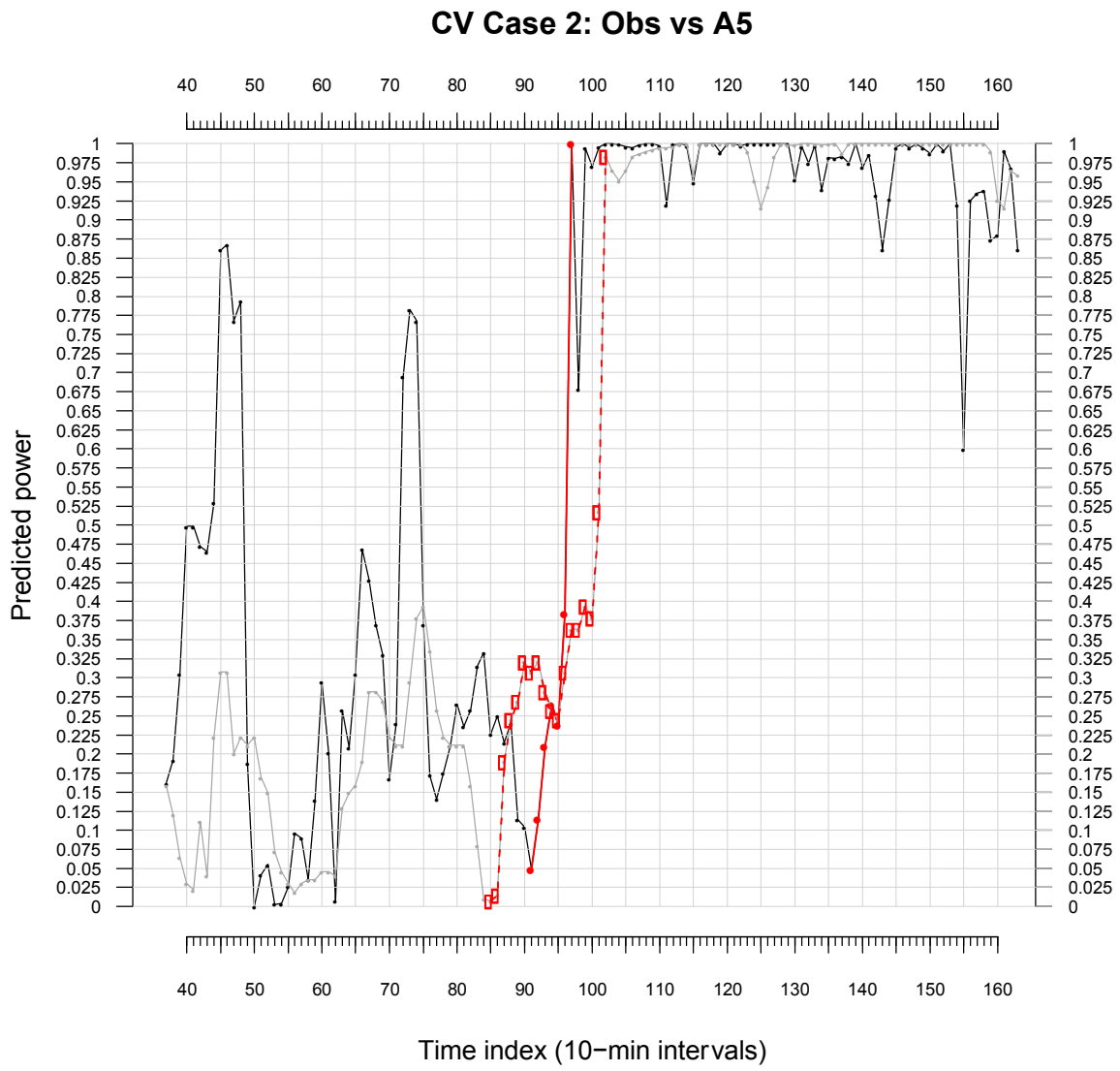

Figure 1. An example of a successful ramp prediction by an ensemble member (MA ensemble member #5 in Chisholm View case study 2). The solid black and grey lines and points represent the "true" and the ensemble prediction of power, respectively. The red lines and points show an identified ramp in the "true" power series, while the dashed red/grey lines and grey-filled red points represent the ramp predicted in the ensemble member power series.

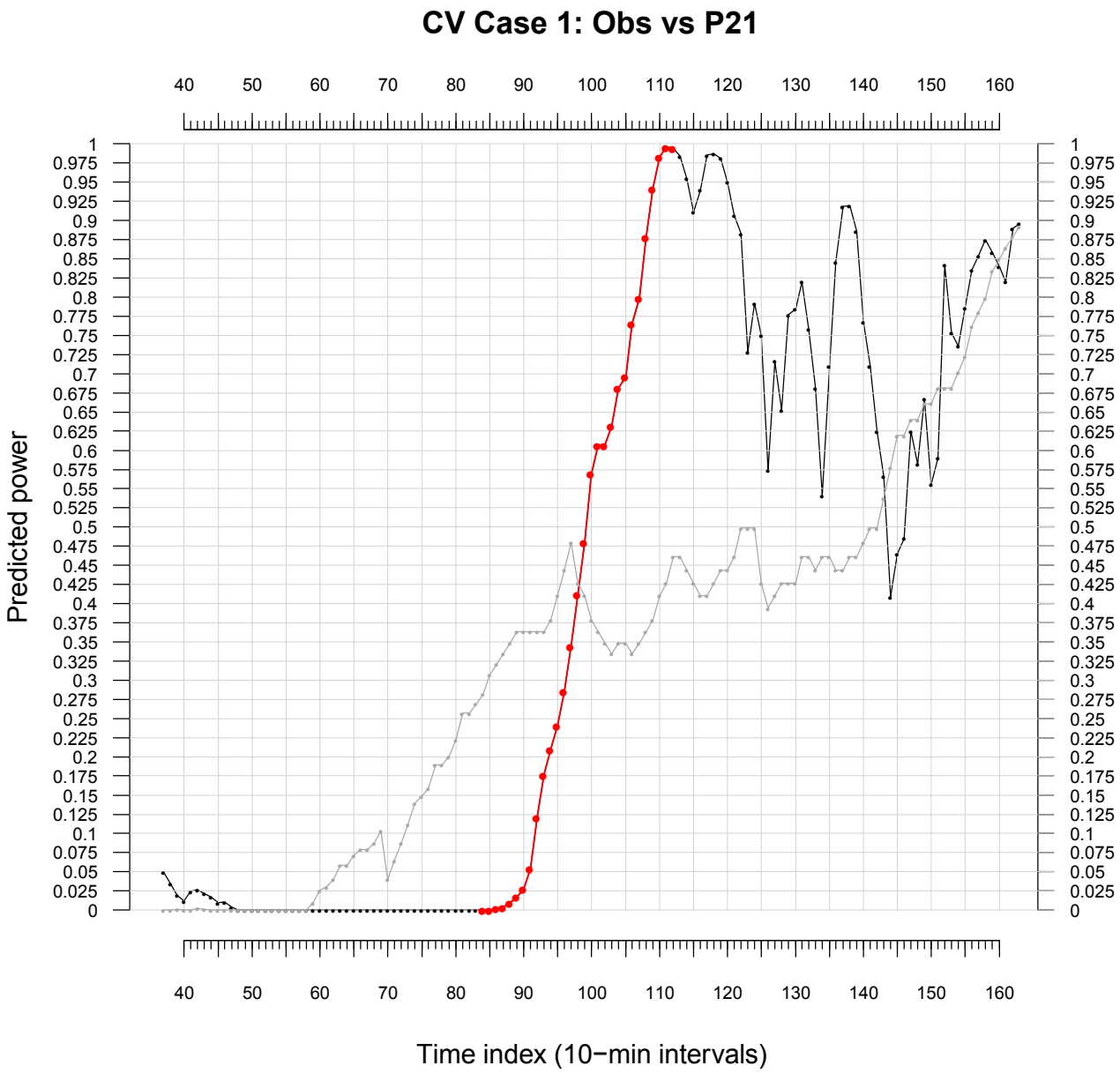

Figure 2. An example of a failed ramp prediction by an ensemble member (MP ensemble member #21 in Chisholm View case study 1). The solid black and grey lines and points represent the "true" and the ensemble prediction of power, respectively. The red lines and points show an identified ramp in the "true" power series. No ramp was predicted by the ensemble member.

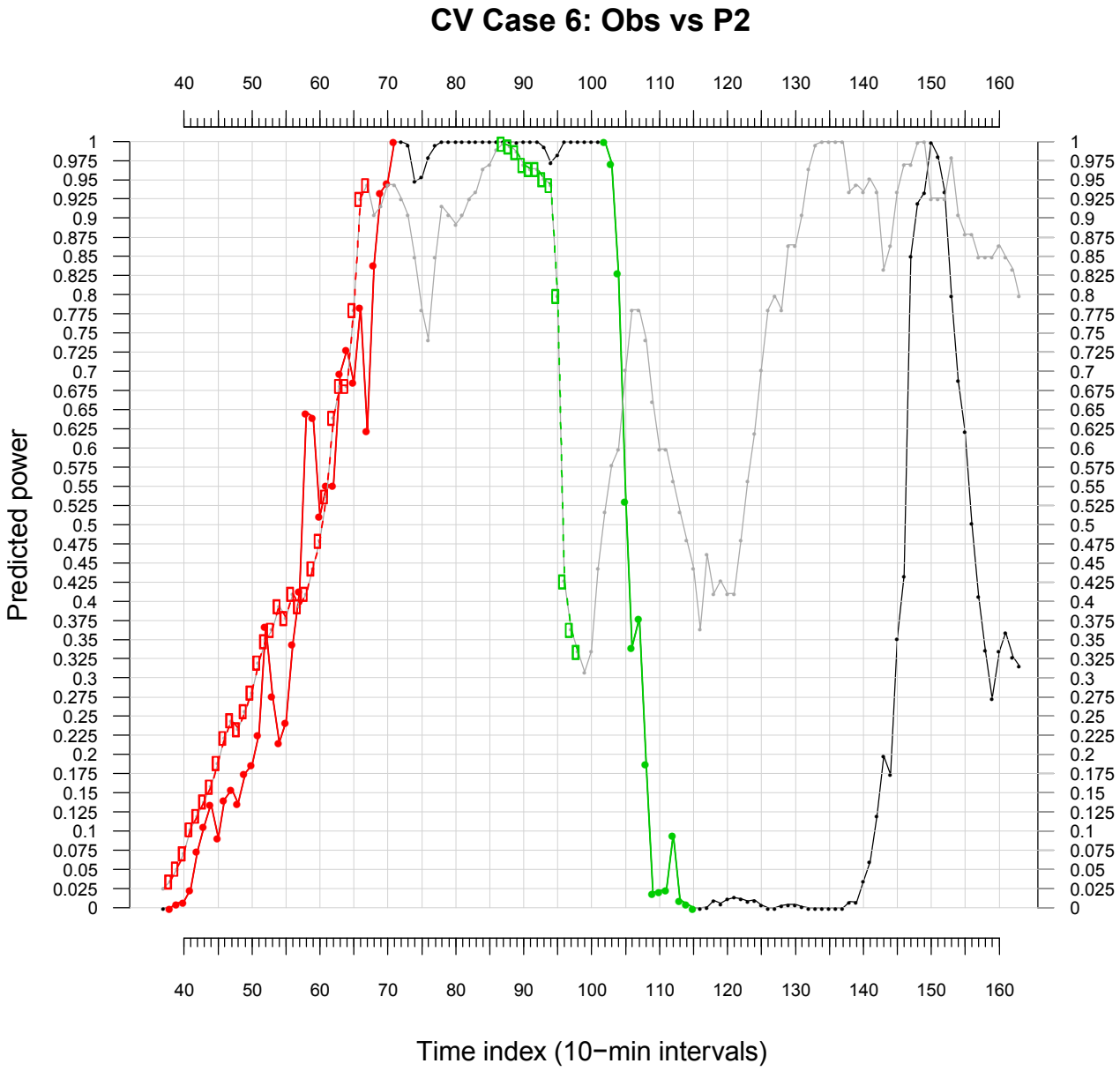

Figure 3. An example of a moderately successful ramp prediction (MP ensemble member #2 in Chisholm View case study 6). The solid black and grey lines and points represent the "true" and the ensemble prediction of power, respectively. The red and green lines and points show two identified ramps in the "true" power series. The dashed red/grey and green/grey lines and grey-filled red and green points represent the two ramps predicted in the ensemble member power series. While the first ramp (in red) was predicted with a high degree of accuracy, the prediction for the second ramp (in green) was predicted less accurately, in terms of the onset and end times, as well as magnitude.

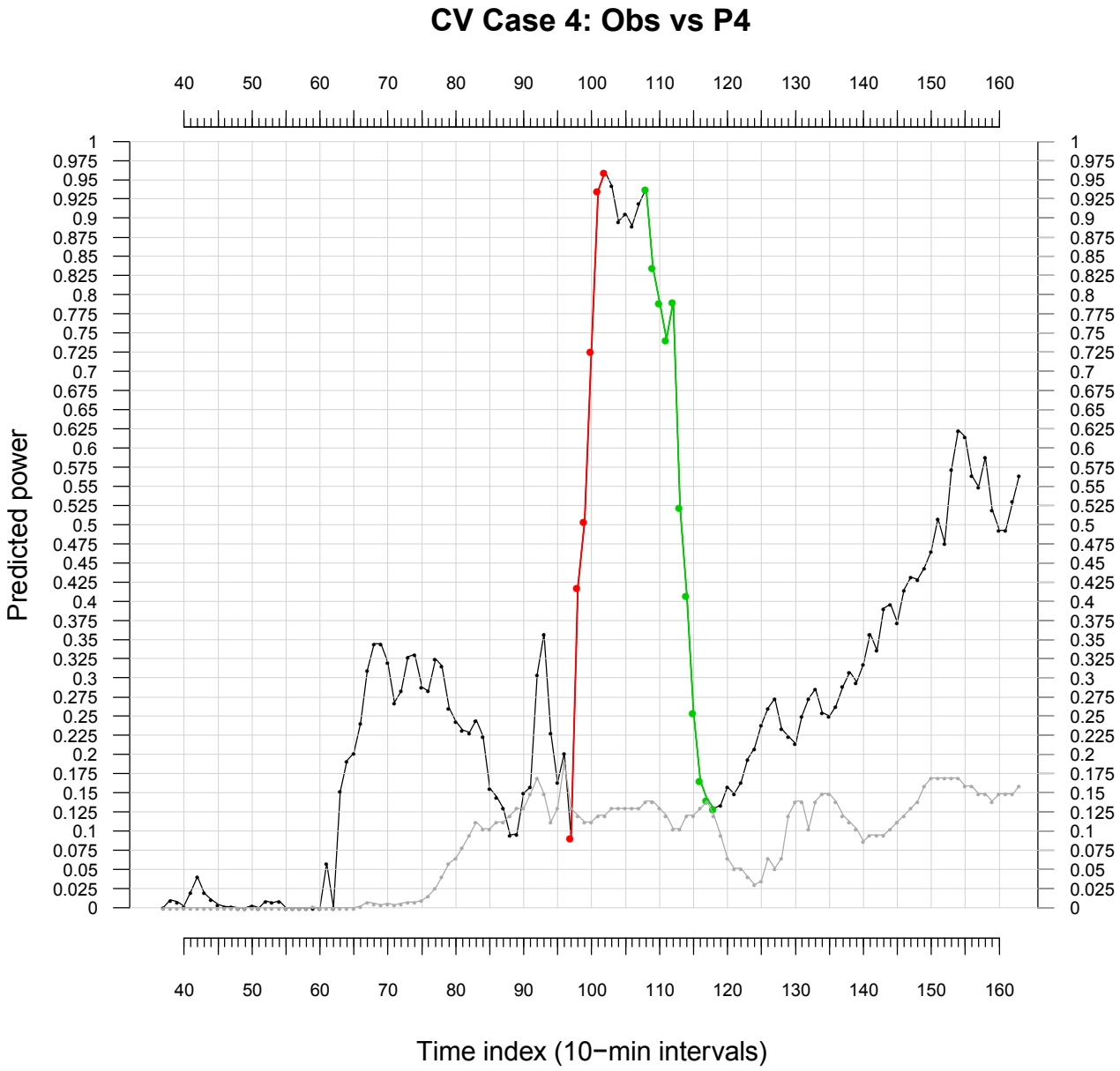

Figure 4. An example of a failed ramp prediction (MP ensemble member #4 in Chisholm View case study 4). The solid black and grey lines and points represent the "true" and the ensemble prediction of power, respectively. The red and green lines and points show two identified ramps in the "true" power series. None of the two ramps was predicted by the ensemble member.