

ITフォーラム 2014

気象庁XMLと他のオープンデータとの 結合／可視化

～気象庁XMLを利用した需要予測の可能性～

2014/02/04

クラウド・テクノロジー活用部会 サブリーダー

菅井康之

(株式会社イーグル)

気象庁XMLによって何が出来るようになるか？

- 従来では、気象データを入手するには有料の配信サービスを利用するか、WebのHTMLをスクレイピングして情報を得る必要があった
 - スクレイピングはサイトの規約に抵触する可能性も
- 気象庁XMLによって、気象データを手軽に入手出来るようになったが、何に活用が出来るか？
 - 天気予報などの気象データを利用したサービスは既に多く展開されている

気象データ単独ではなく、他のデータを
組み合わせることで何か出来ないかを考える

- 予測データ
 - 天気予報、週間天気予報、季節予報、1か月・3か月予報、
暖・寒候期予報・・・
- 観測データ
 - 生物季節観測、紫外線観測、火山観測・・・
- 速報データ・その他
 - 地震速報、津波警報・注意報・予報、気象警報・注意報、
竜巻注意報、記録的短時間大雨情報、台風情報・・・
- 天候などのデータに関しては予測データが主であり、実測値は含まれていない
- 予測データを使って何か予測出来ないか？

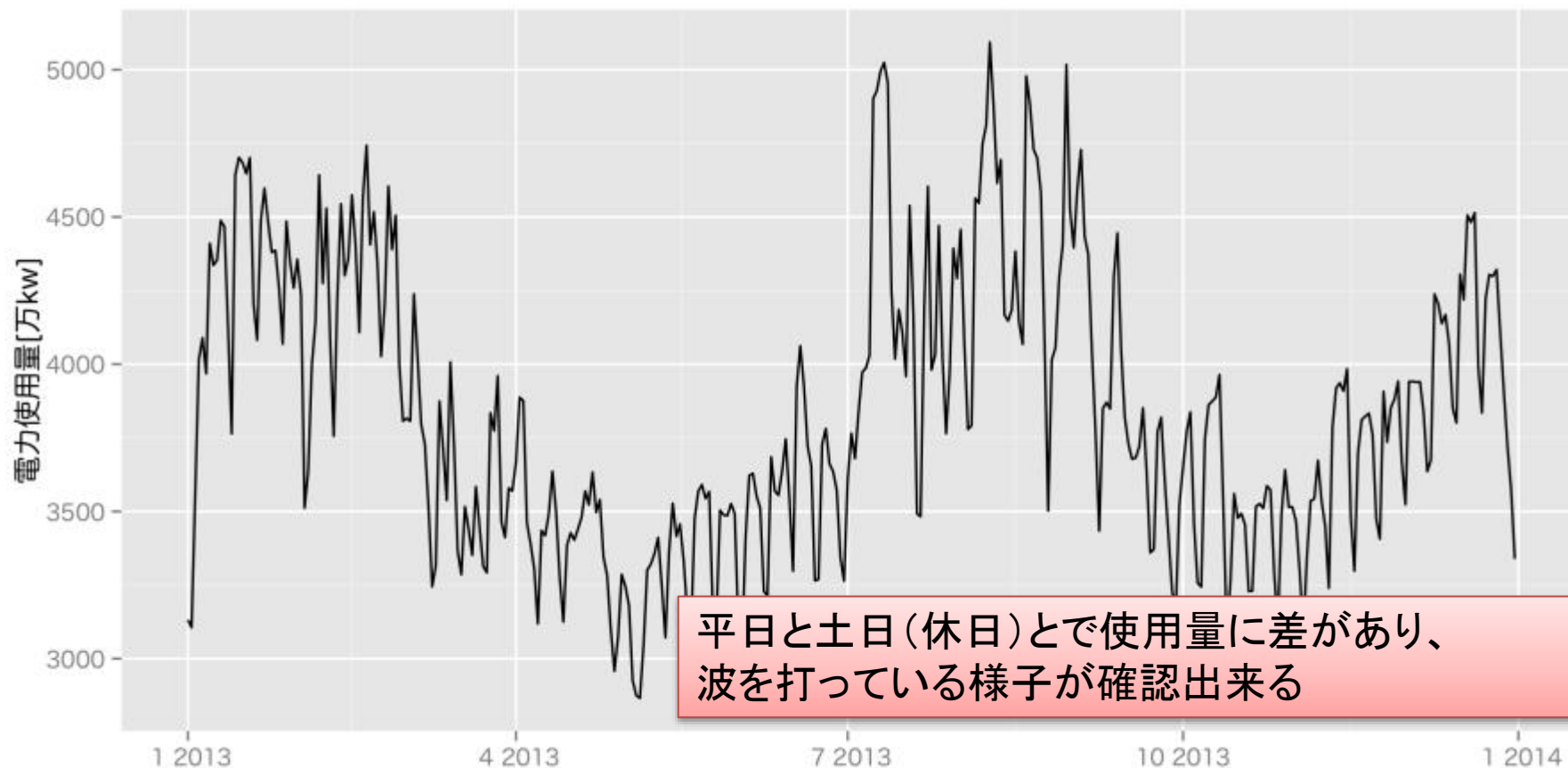
- 天候、降水確率、最高気温、最低気温といった気象データと相関のあるデータについて、気象庁XMLから予測モデルを構築出来ないか？
- 数日先の予測では、週間天気予報のデータが利用出来る
- 天候や気温と相関のあるデータ・・・
 - 電力の使用量と最高気温に相関があるという説は良く耳にする

データを可視化しながら、最高気温を基に
電力使用量についての需要予測が行えるかを検証する

本資料で利用したデータについて

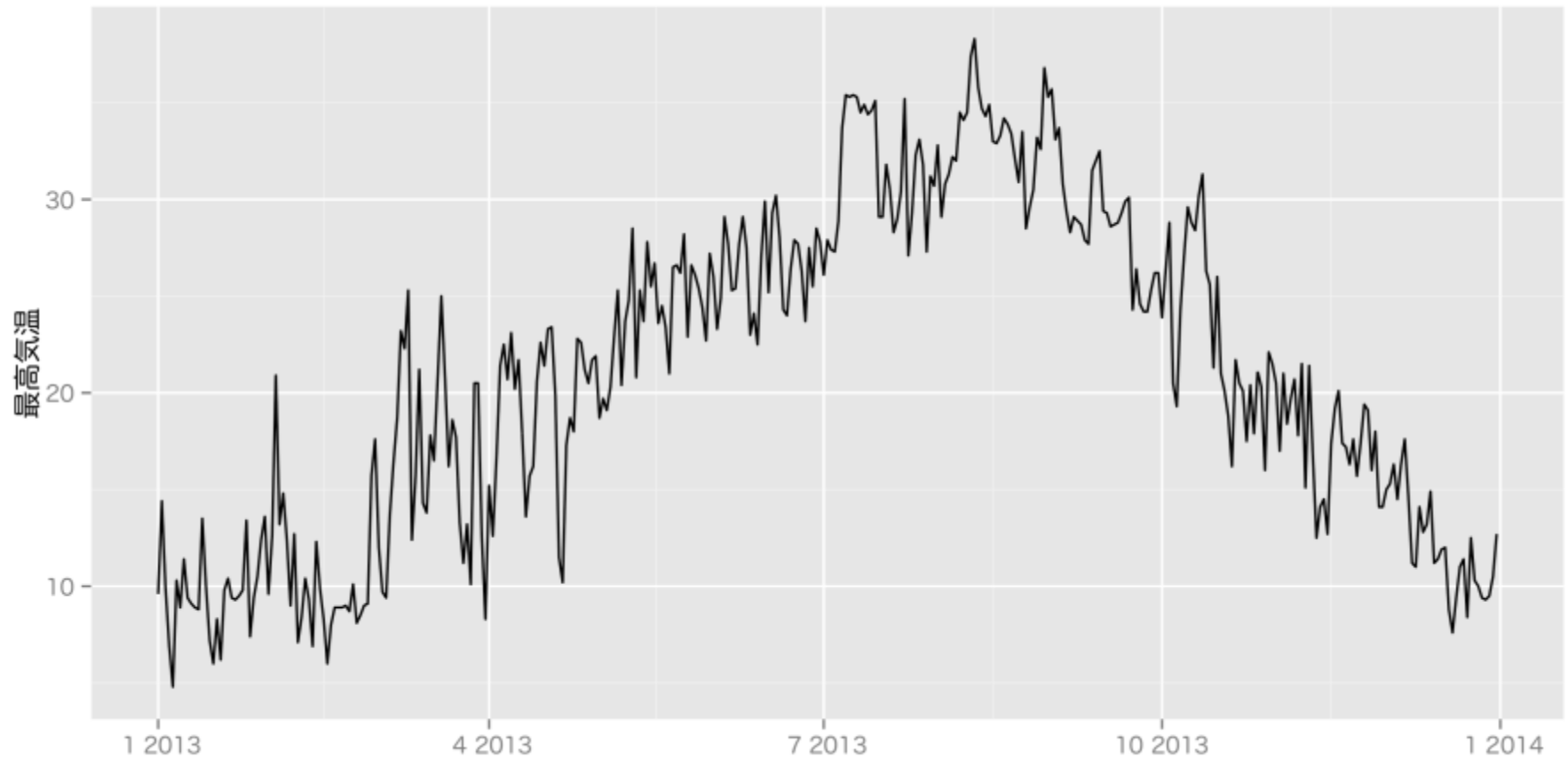
- 予測モデルは、電力使用量と最高気温の実測値を基に構築する
- 電力使用量については、東京電力が過去の使用量を公開しているため、2013年のデータをダウンロードして使用
 - <http://www.tepco.co.jp/forecast/html/download-j.html>
 - CSV形式
- 気温については、気象庁が過去の統計情報を公開しているため、2013年のデータをダウンロードして使用
 - <http://www.data.jma.go.jp/gmd/risk/obsdl/index.php>
 - CSV形式

2013年の電力使用量（ピーク時）



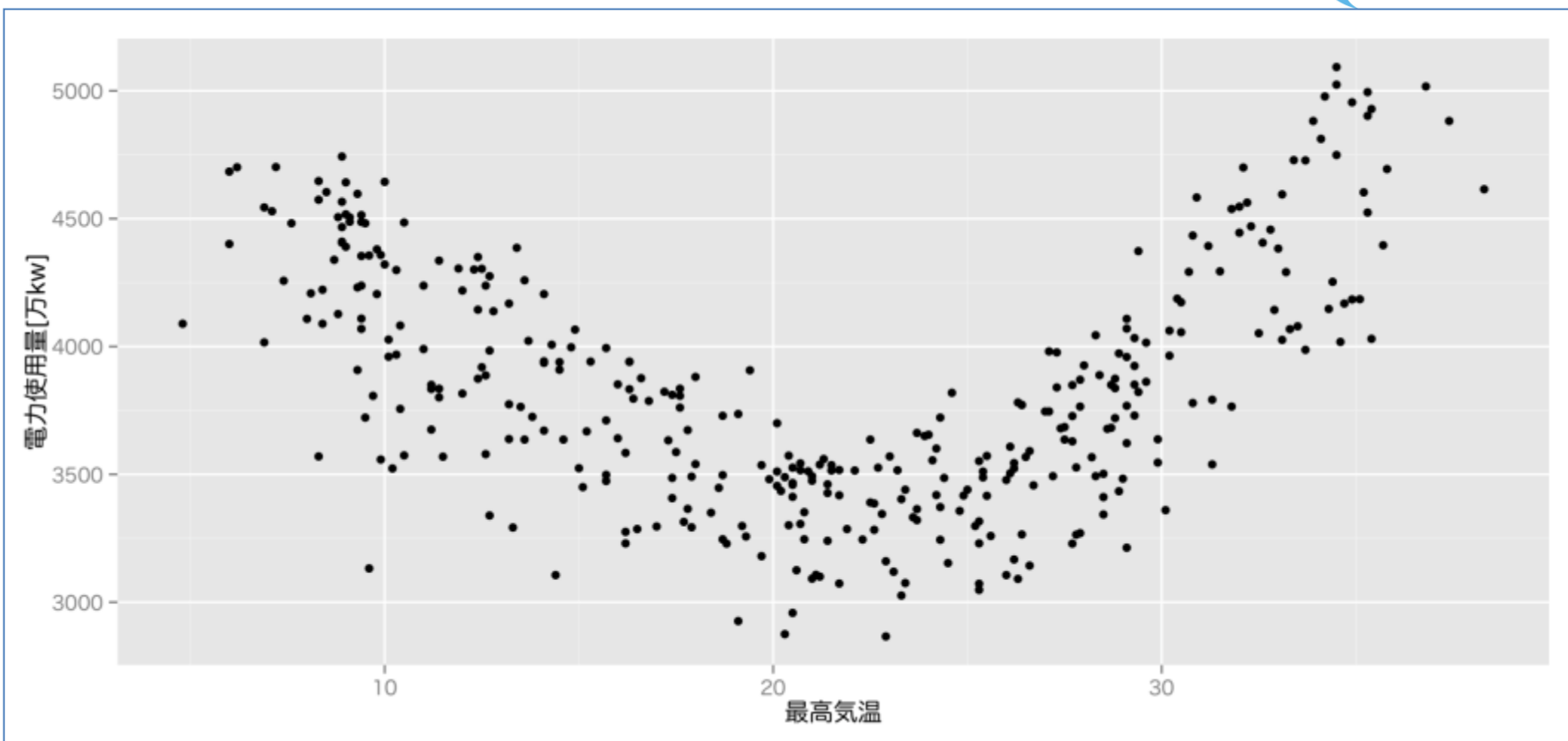
1時間辺りの電力使用量から、1日の最大電力使用量（ピーク）を算出

2013年の最高気温



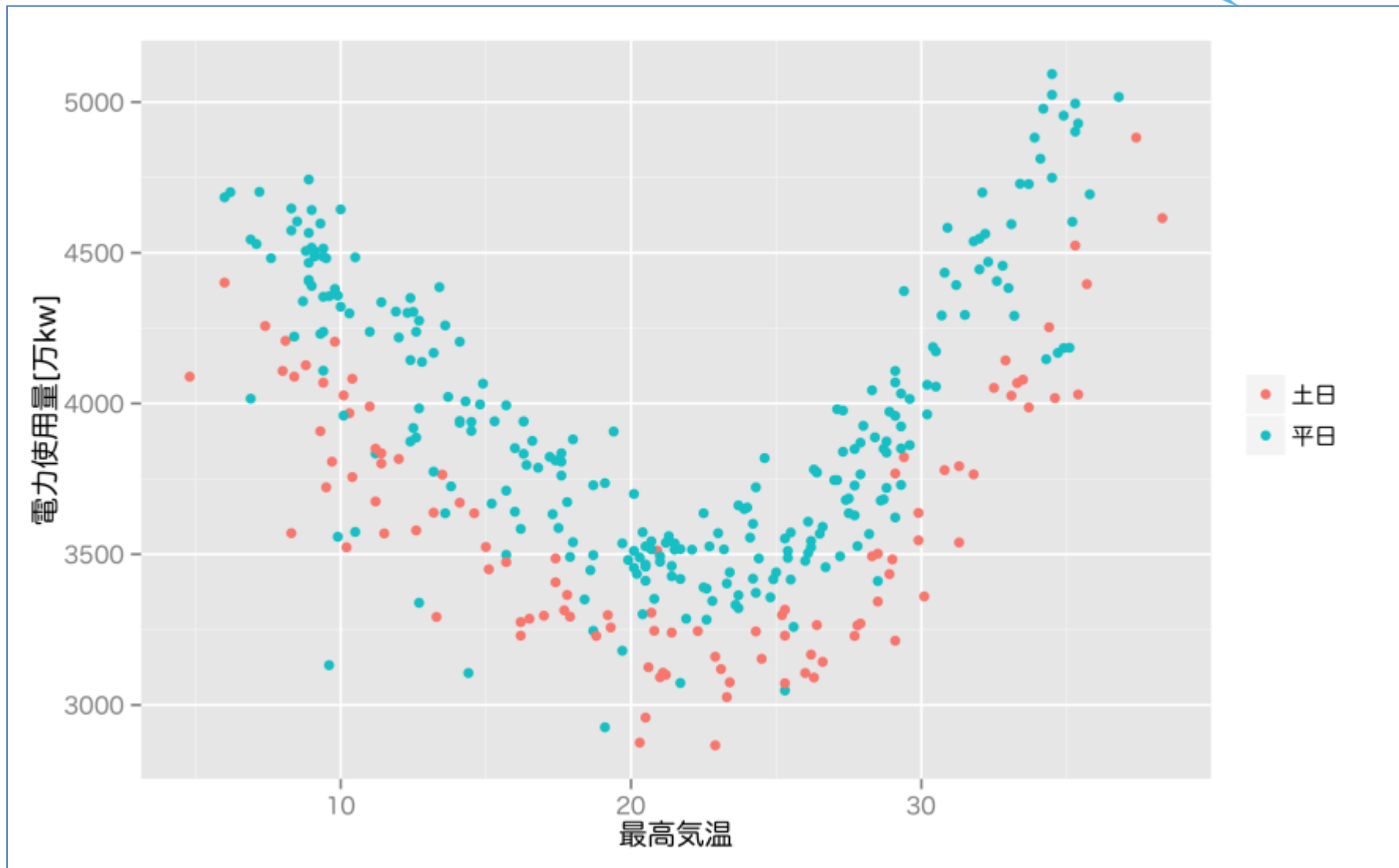
季節の中でも、最高気温にバラつきがあることが確認できる

電力使用量と最高気温の相関



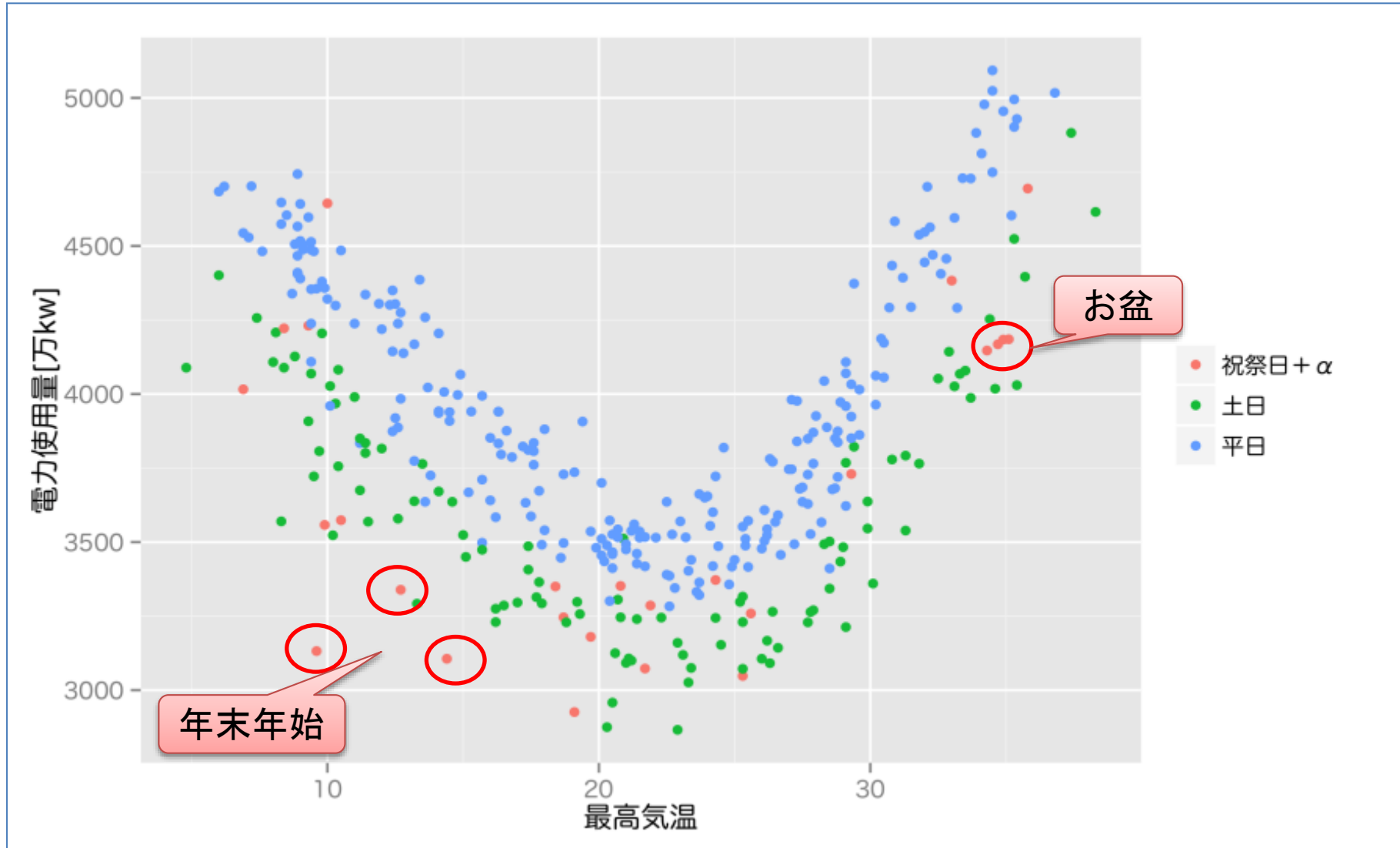
気温が上がるor下がる程、使用量が増加する傾向

電力使用量と最高気温の相関（土日・平日分類）



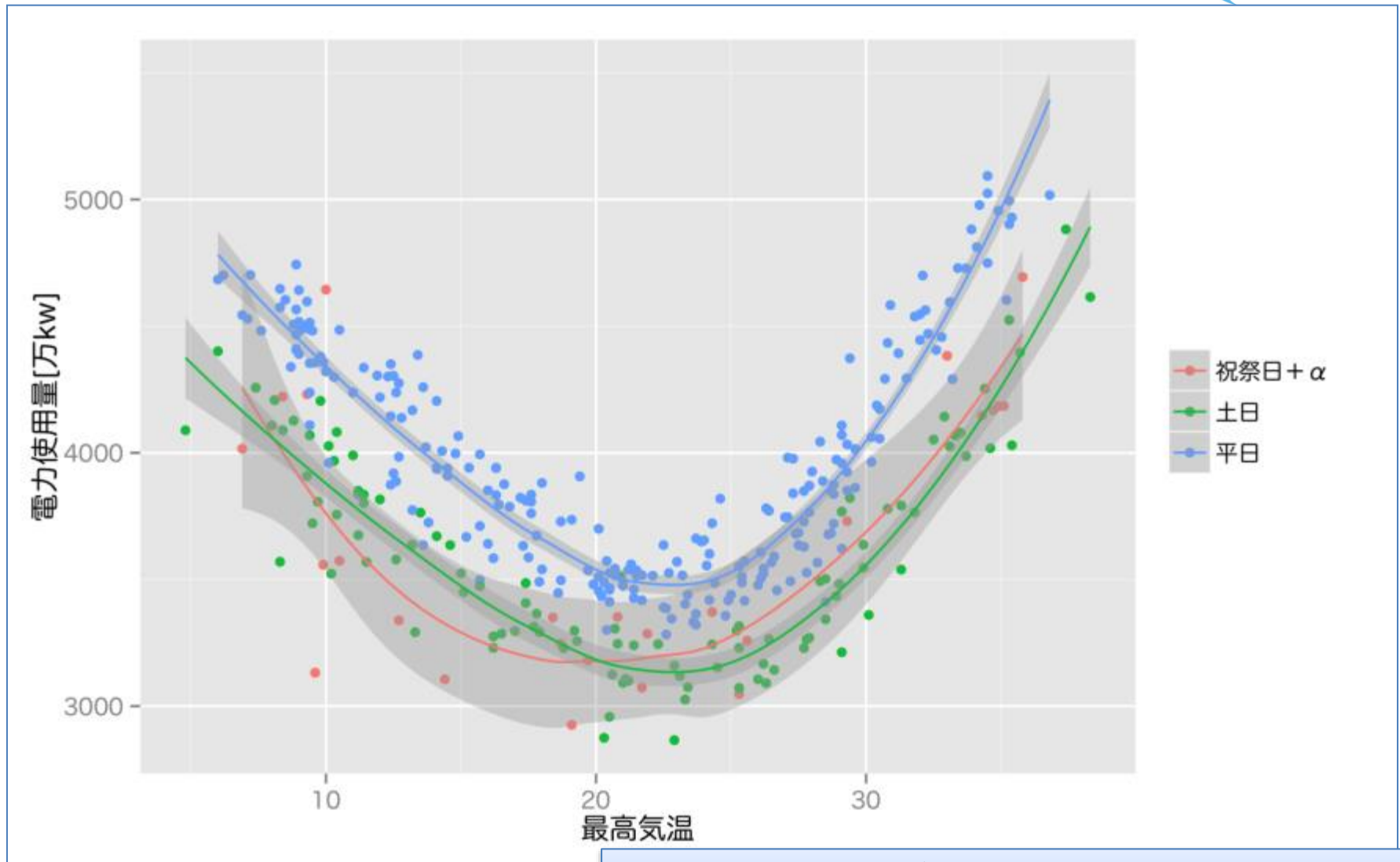
平日と土日とで使用具合が異なるため、分類

電力使用量と最高気温の相関（祝日分類）



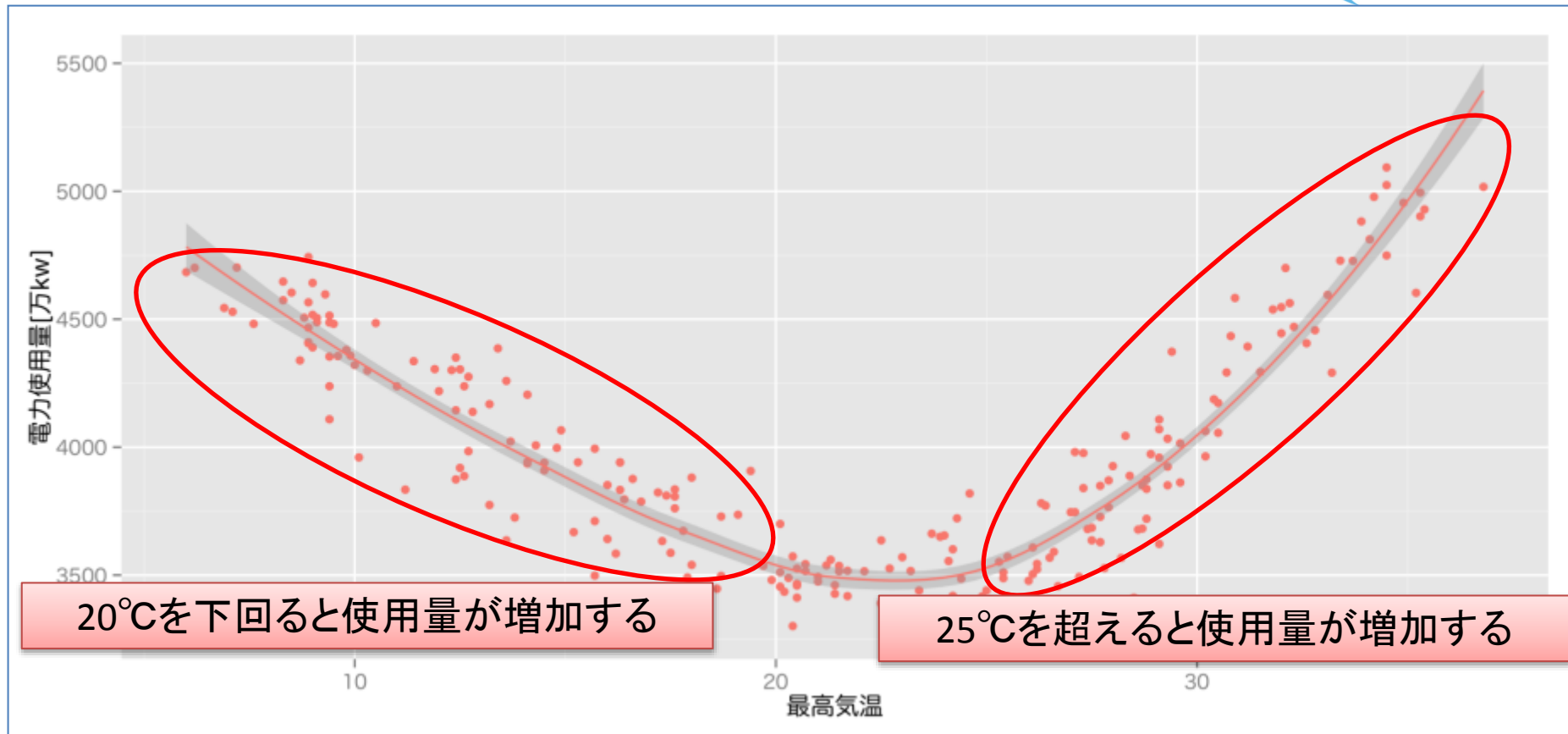
平日、土日に加えて祝祭日やお盆、年末年始をさらに分類

電力使用量と最高気温の相関（近似線）



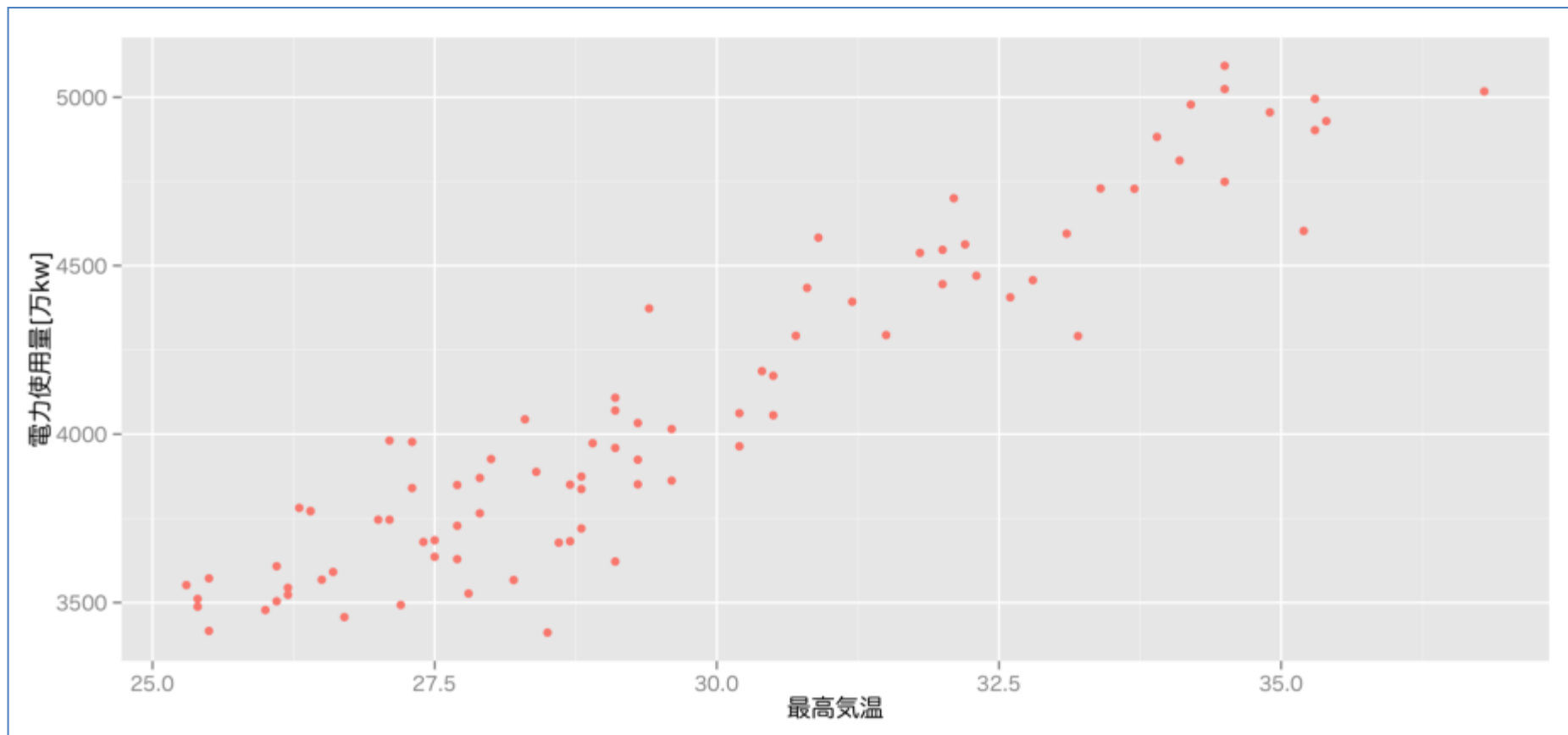
平日と土日とで同様の傾向
平日が予測出来れば、土日も予測可能と考える

電力使用量と最高気温の相関（平日のみ抜粋）



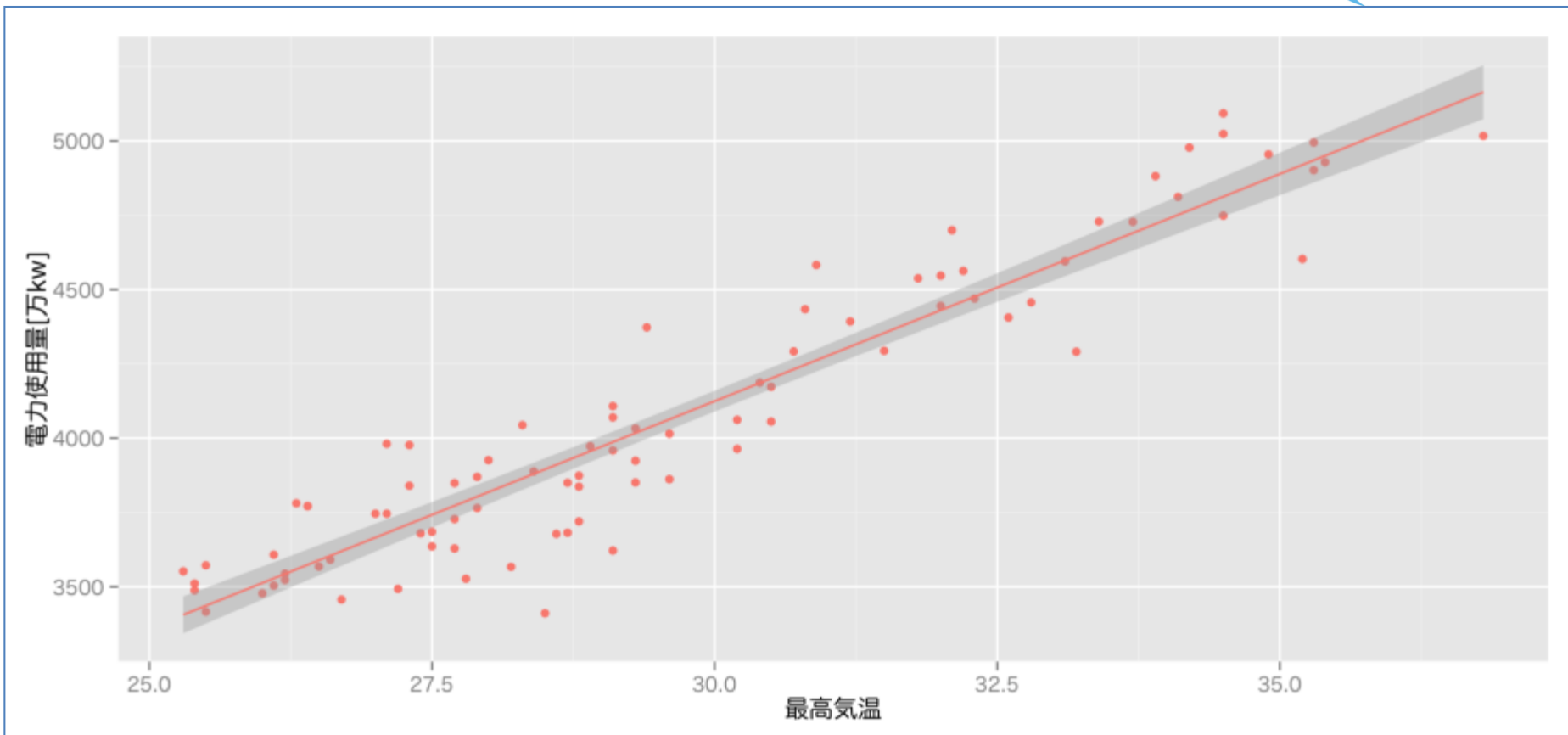
25°Cを超えた場合の予測モデルを考える

電力使用量と最高気温の相関（平日25度超）



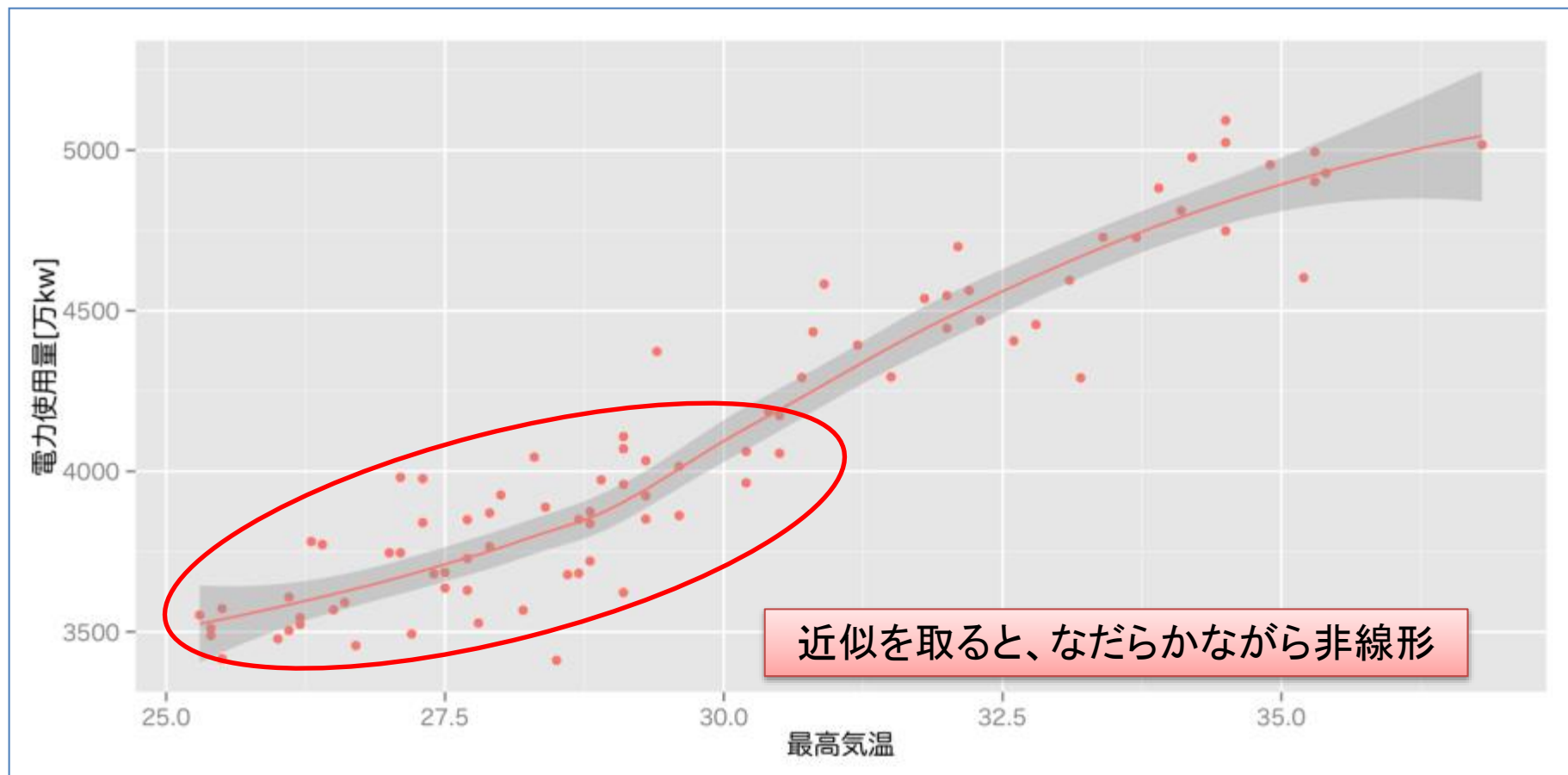
相関係数も0.94を越え、強い相関があるといえる
(最低気温と電力使用量の相関係数は0.89)

電力使用量と最高気温の相関（回帰直線）



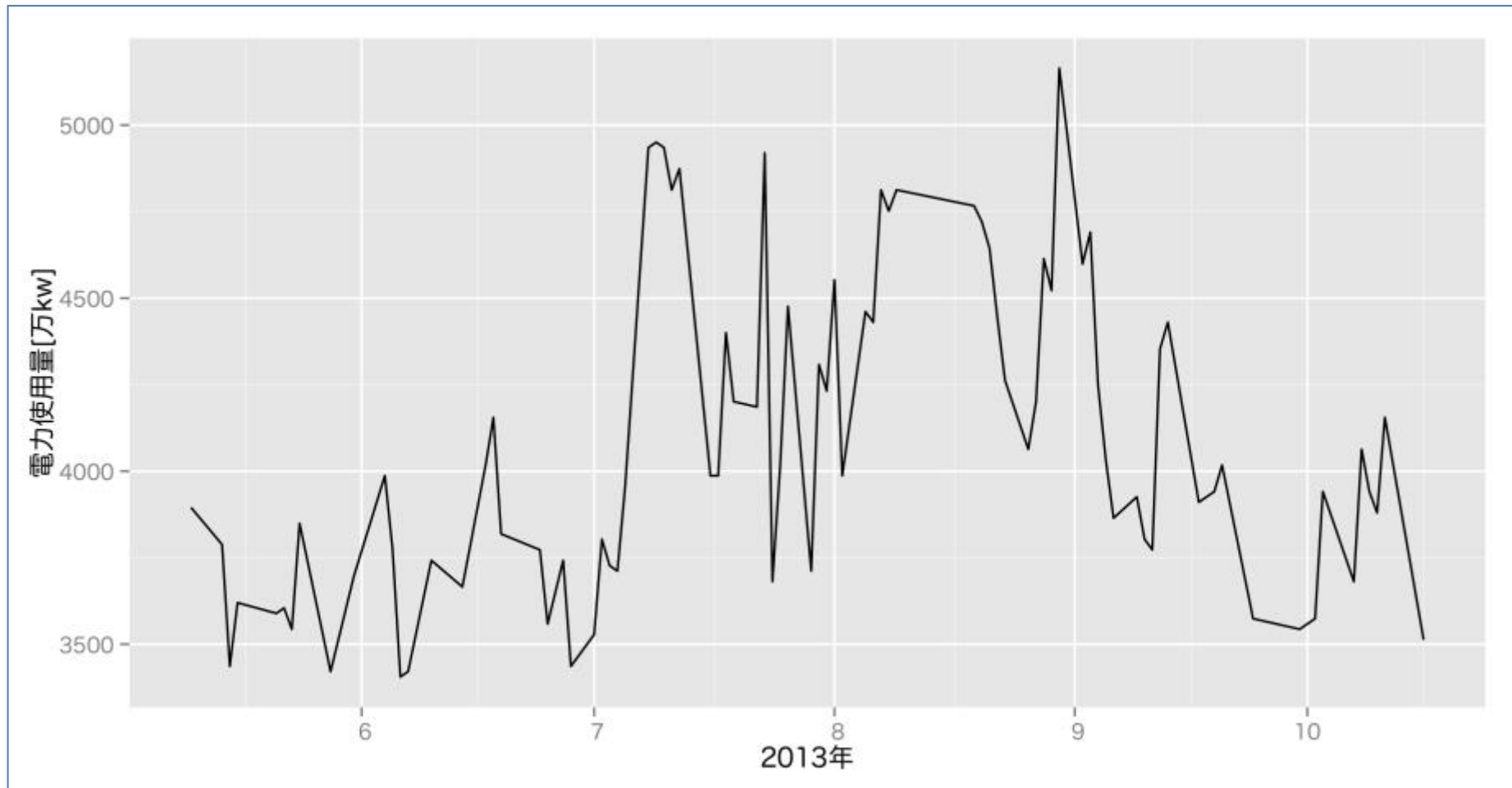
最高気温だけの単回帰分析でいけるか？

電力使用量と最高気温の相関（近似線）



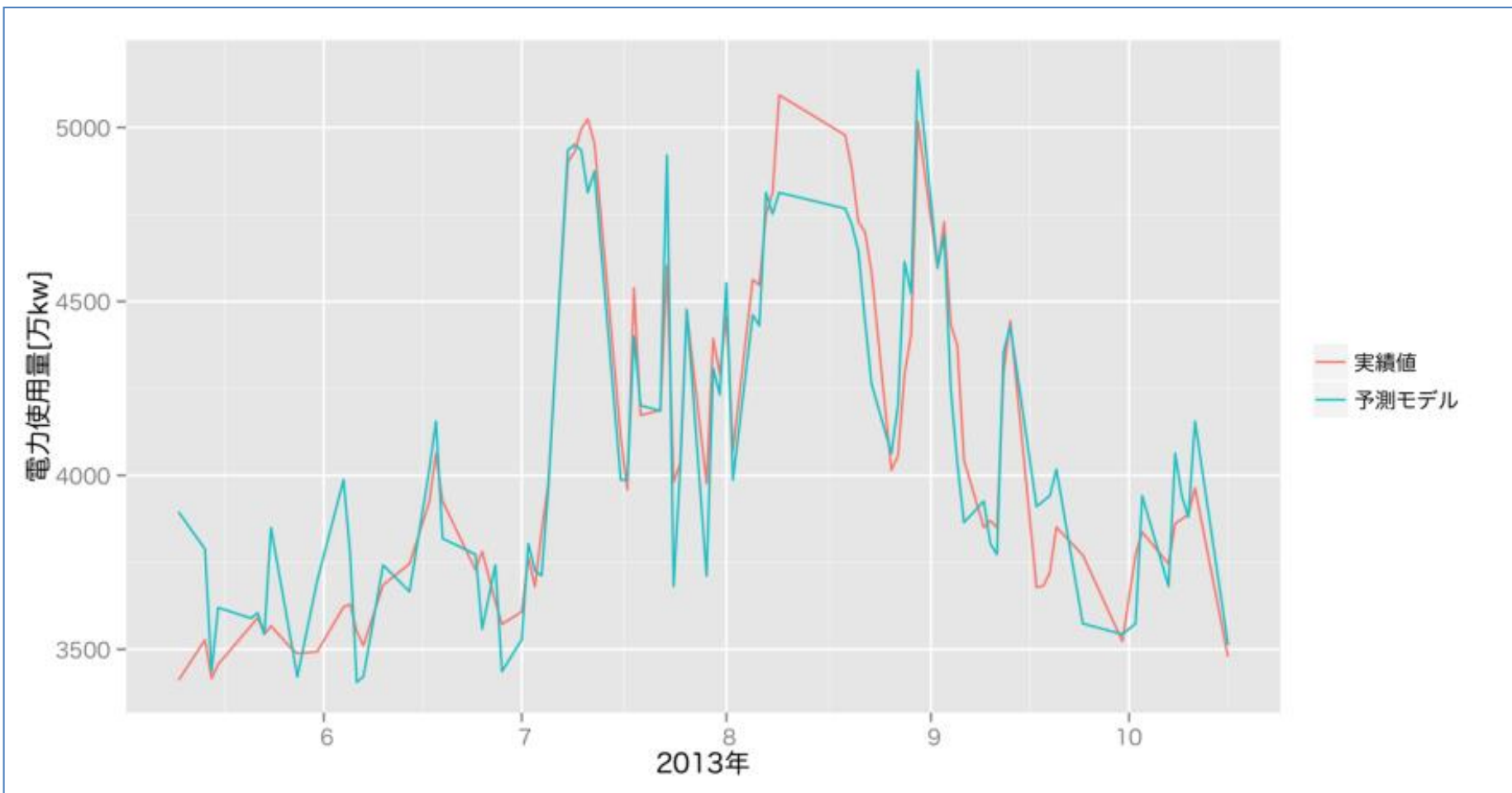
最低気温を使用した重回帰も検討する

単回帰分析（最高気温）による予測モデル



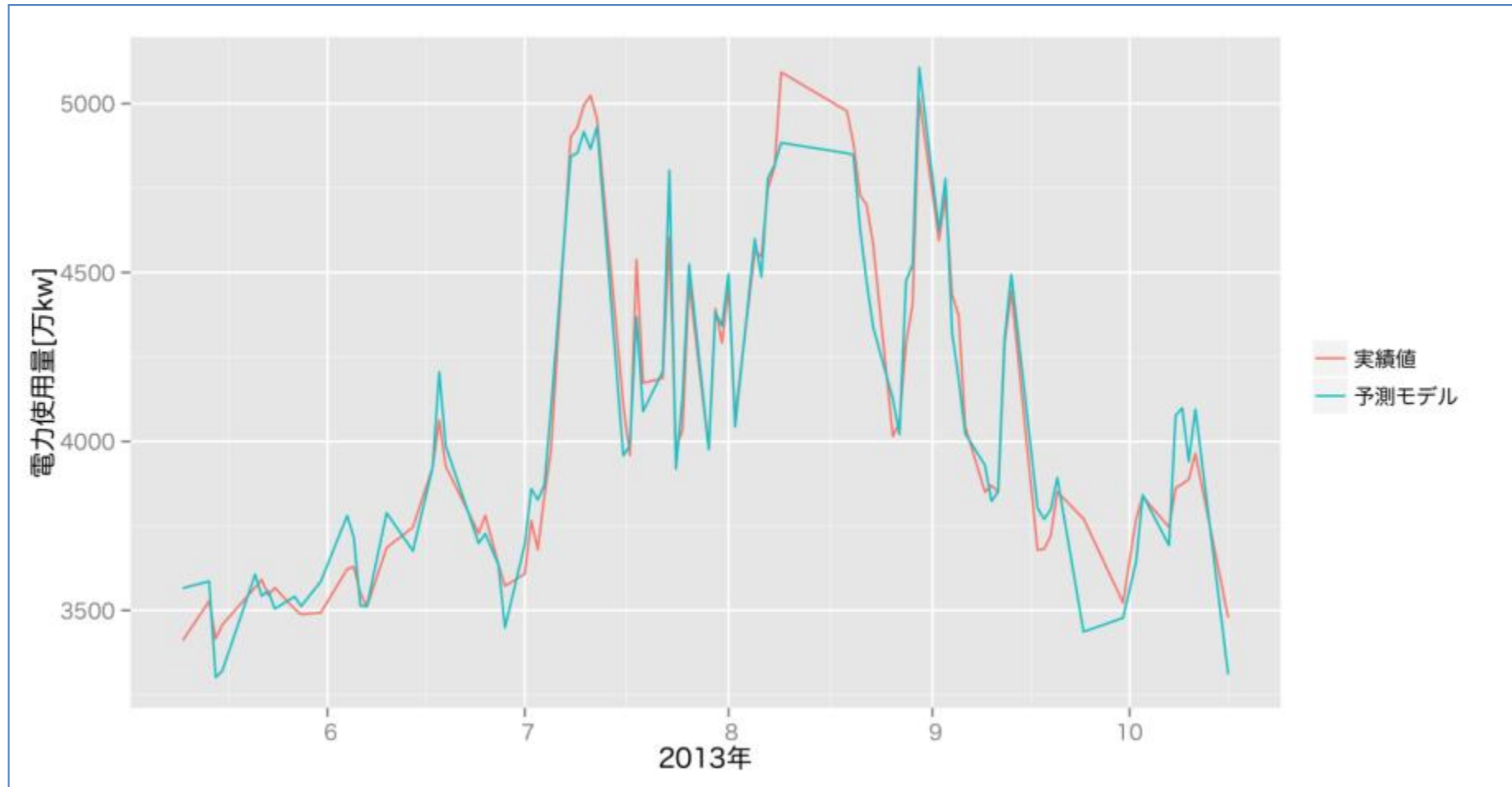
予測モデル: 電力使用量 = $-463.545 + \text{最高気温} \times 152.937$

予測モデルと使用量とを比較

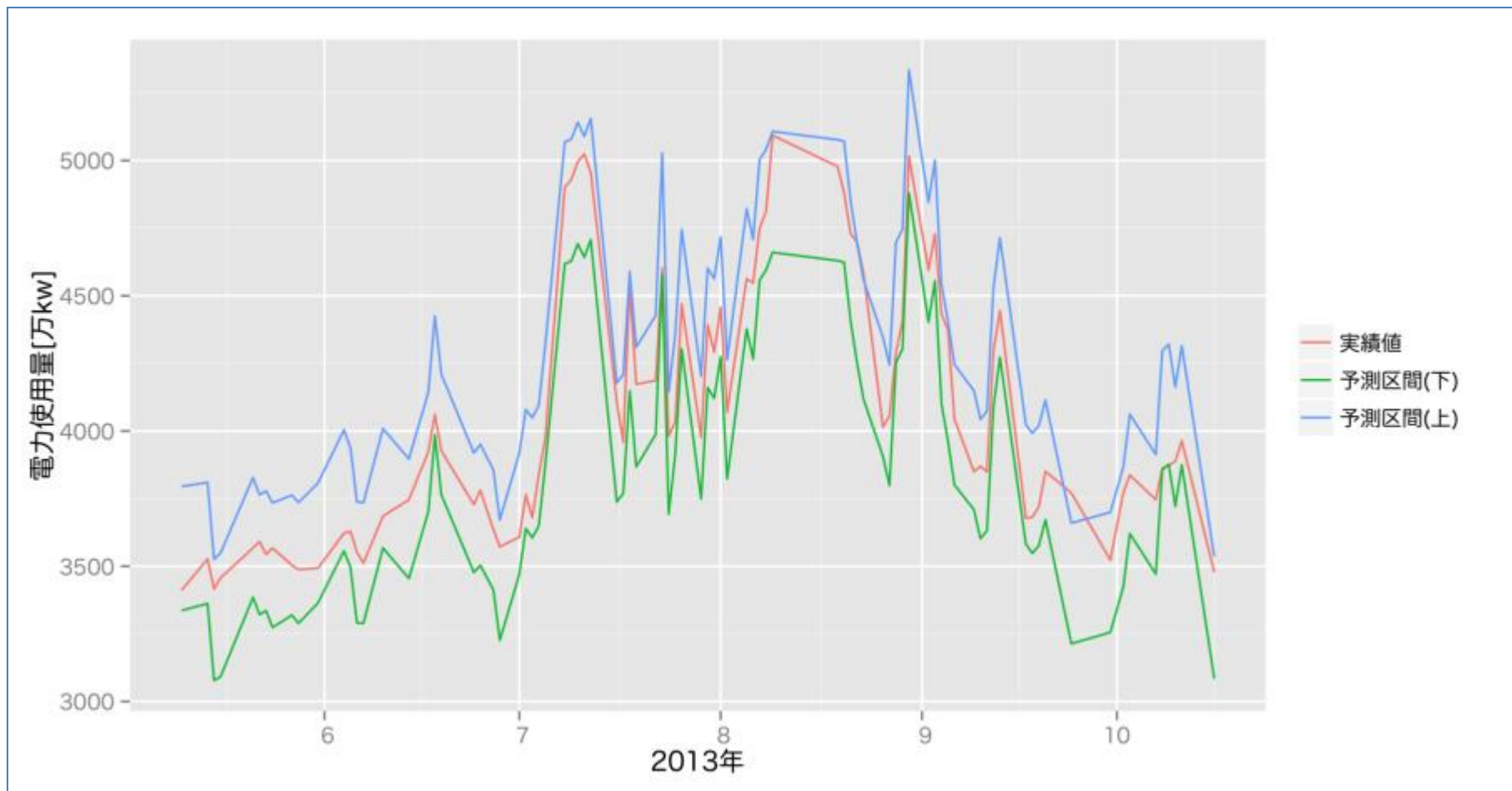


26°C～30°C辺りの精度に難

重回帰分析（最高気温＋最低気温）による予測モデル



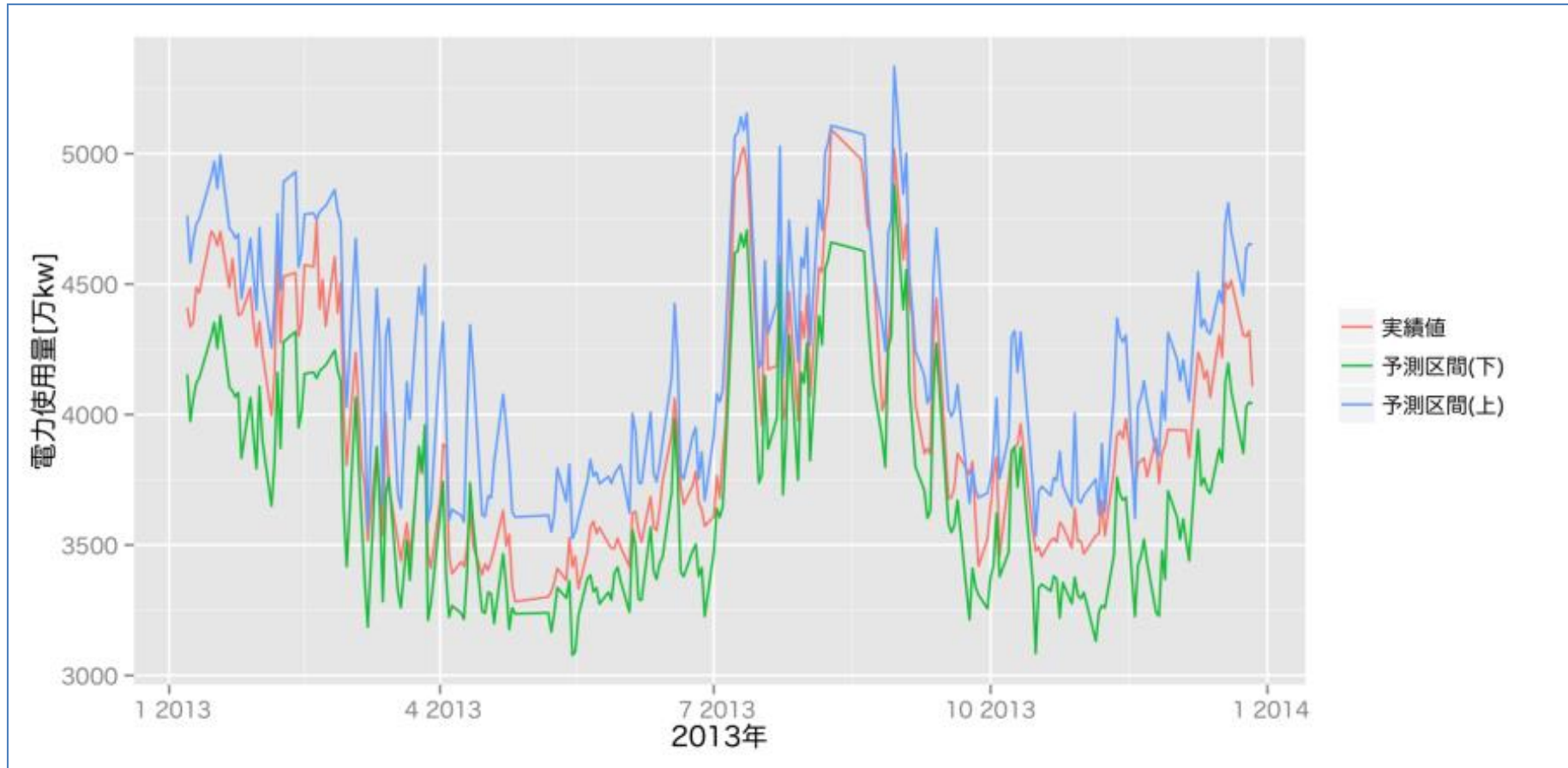
予測モデル: 電力使用量 = $-341.562 + \text{最高気温} \times 102.260 + \text{最低気温} \times 61.284$



使用量のピークの上限值を予測するため、予測モデルから予測範囲を算出

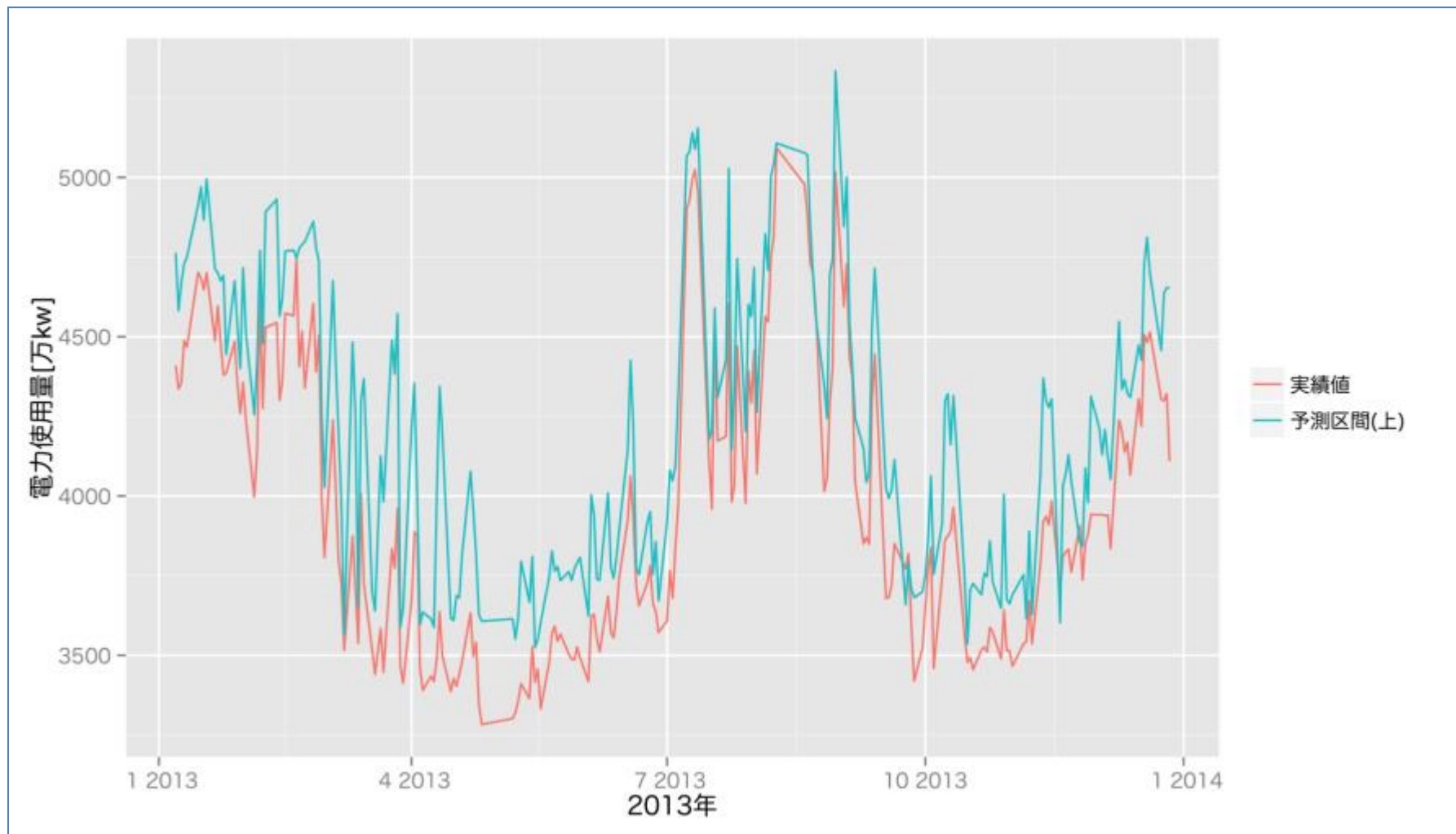
94.2%程の精度

1年間の実測値を基に、予測モデルを構築



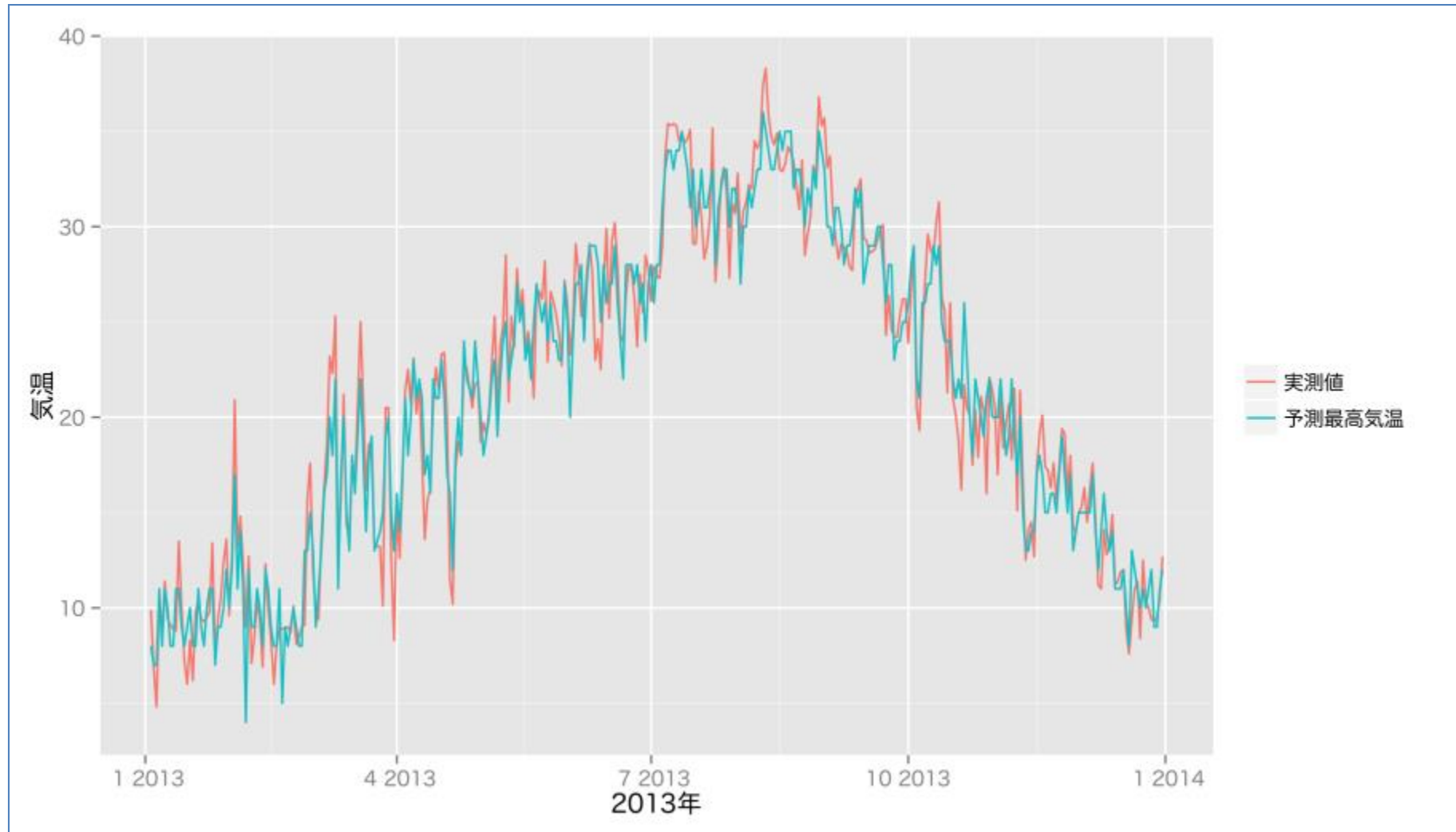
($\sim 20^{\circ}\text{C}$ / $20^{\circ}\text{C} \sim 25^{\circ}\text{C}$ / $25^{\circ}\text{C} \sim$) それぞれ異なる予測モデルを構築

予測区間の上限と実績値

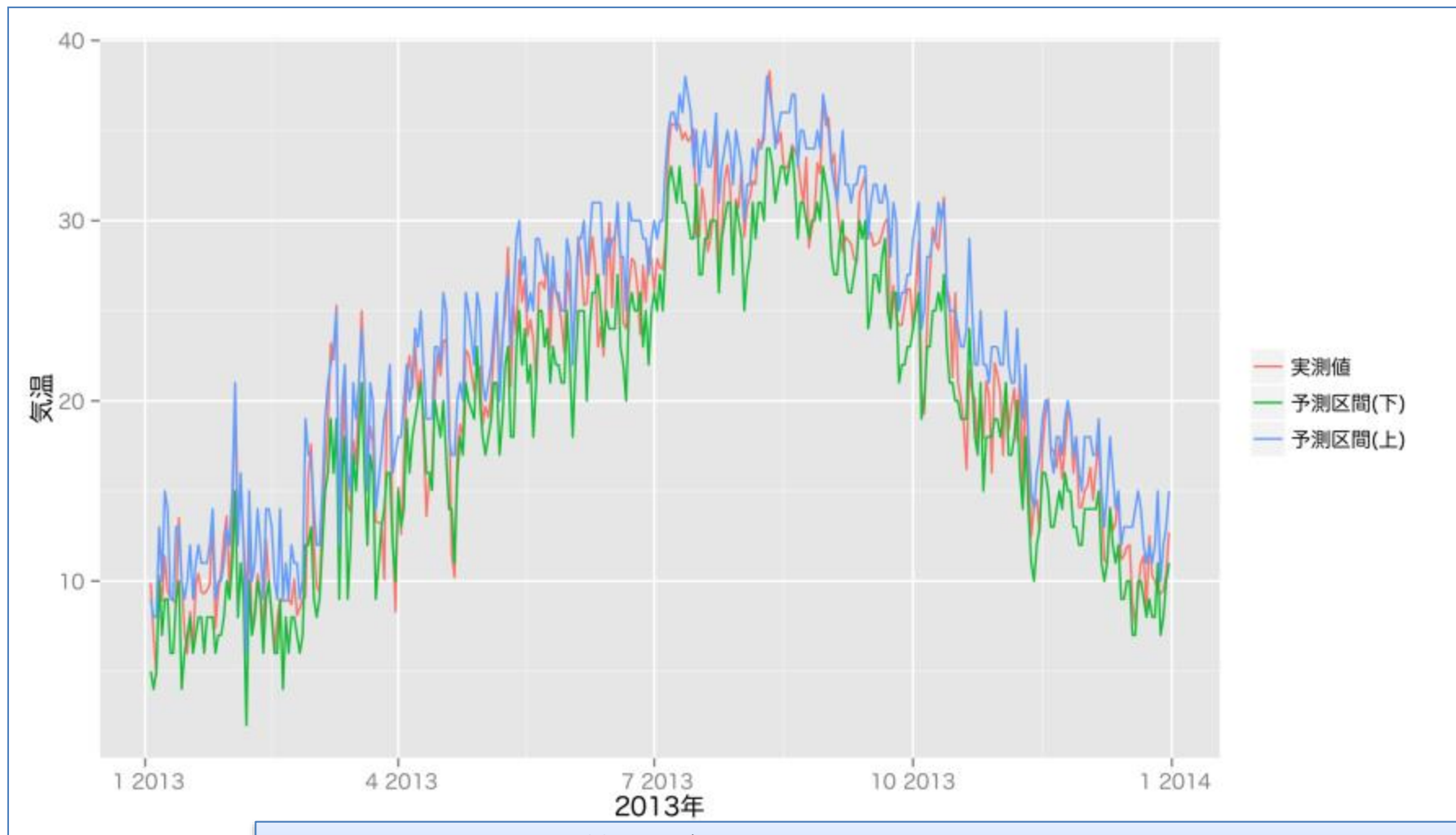


- 最高気温、最低気温を基に、電力使用量の需要を予測することが可能
 - 他のデータを加える事により、より精度を上げられるが最高気温、最低気温だけでここまでの精度を得られている
- ここまでは気象データの実績値（CSV）を使用していたが、気象庁XMLで取得できる気温データがどの程度の精度かを確認する
 - 何日後までなら予測出来そうか

2日後の予測最高気温と実測値

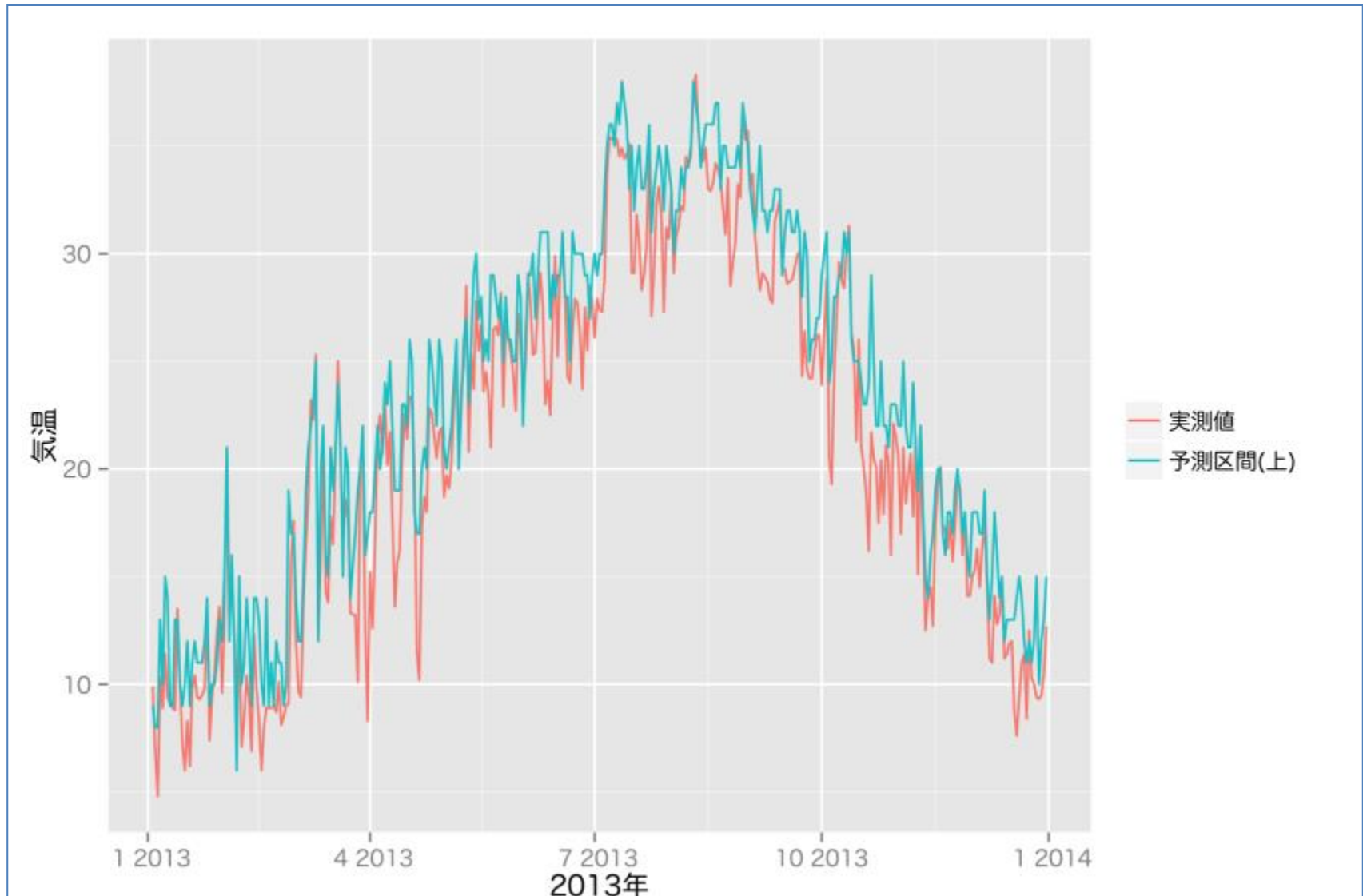


2日後の予測最高気温と実測値



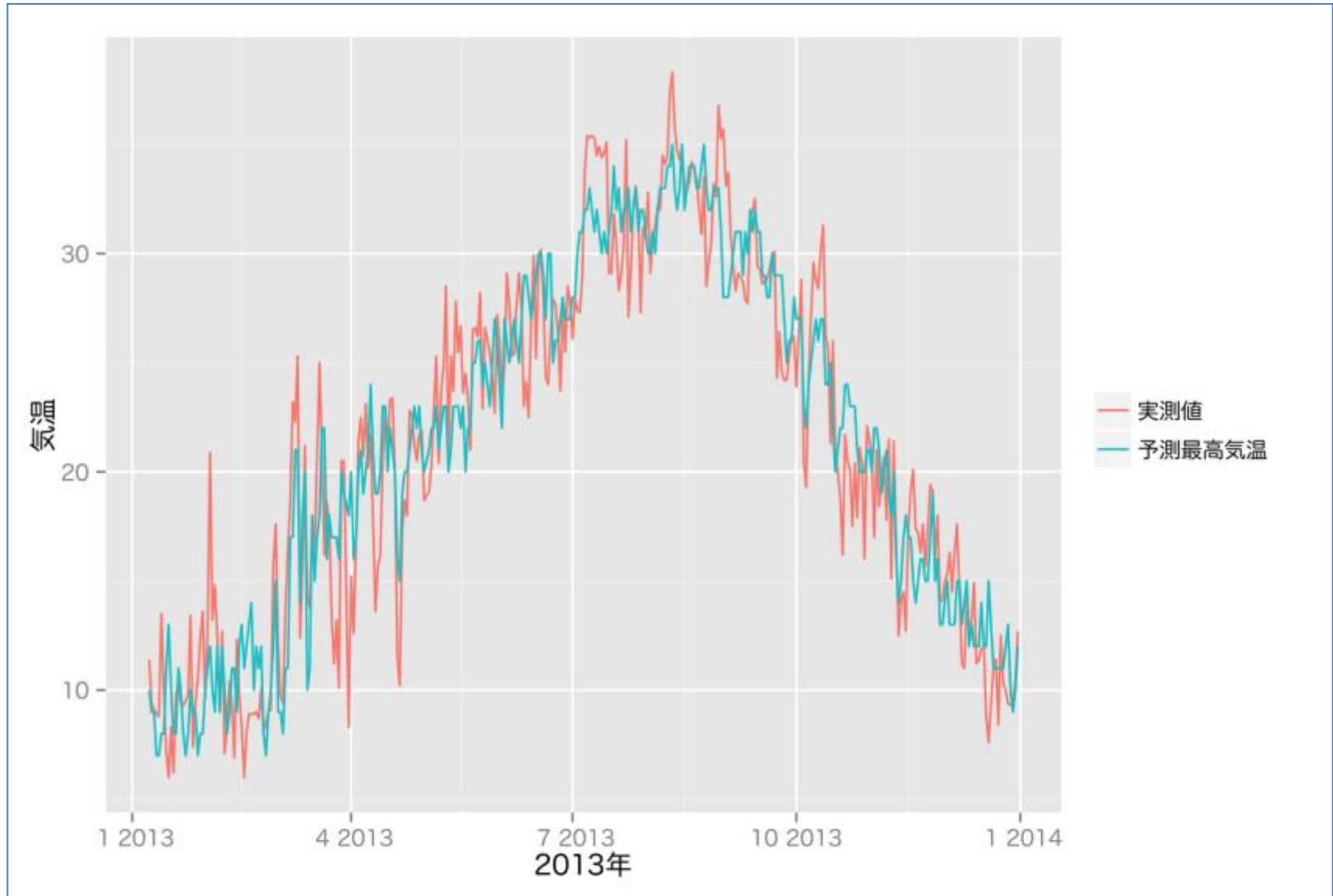
予測気温には予測範囲があり、
「実況の気温がこの範囲に入る確立はおよそ80%である」とのこと
また降水の有無については信頼度という指標もある←今回は考慮しない

2日後の予測最高気温と実測値

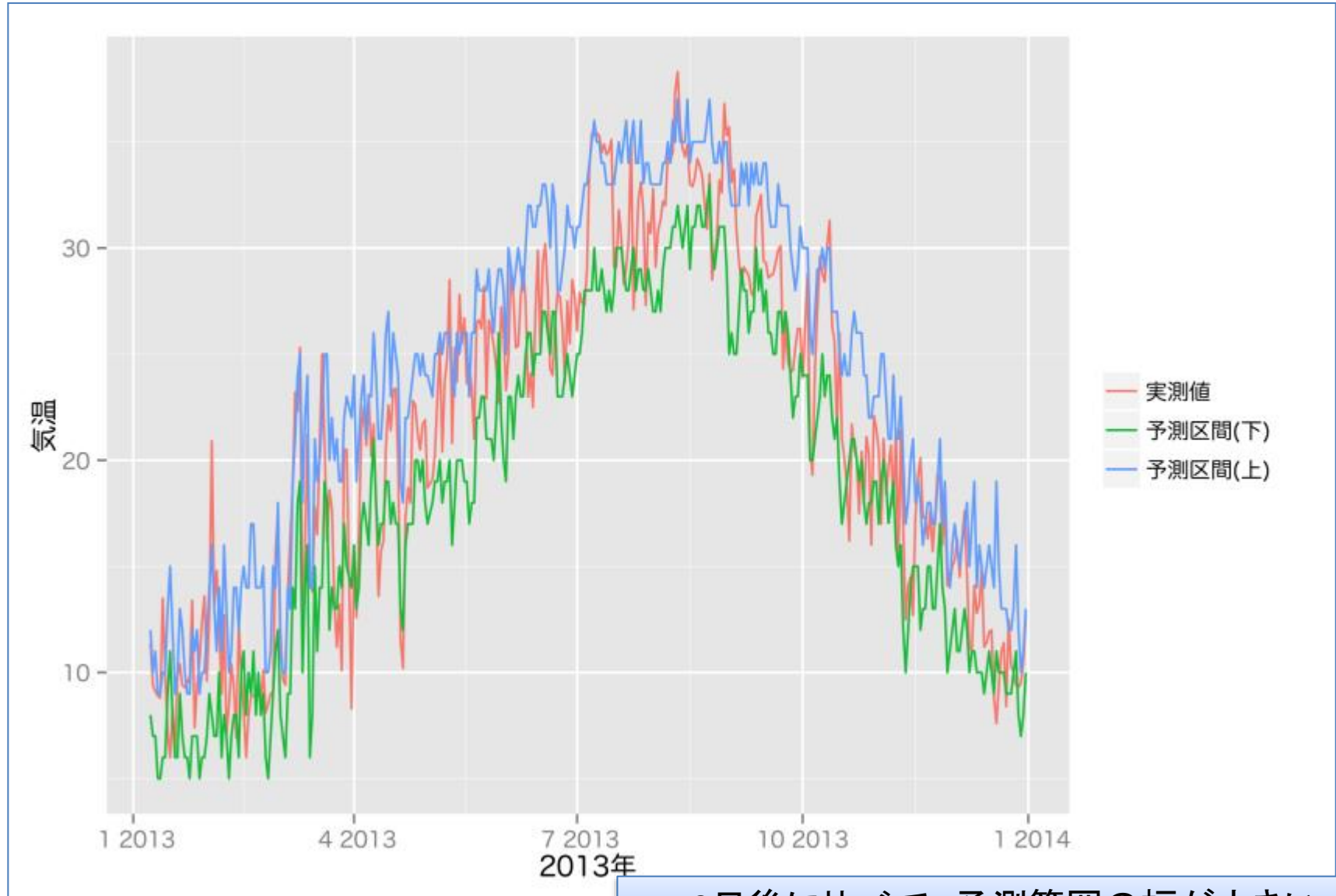


予測範囲の上端を最高気温の予測最大値とみなす

7日後の予測最高気温と実測値

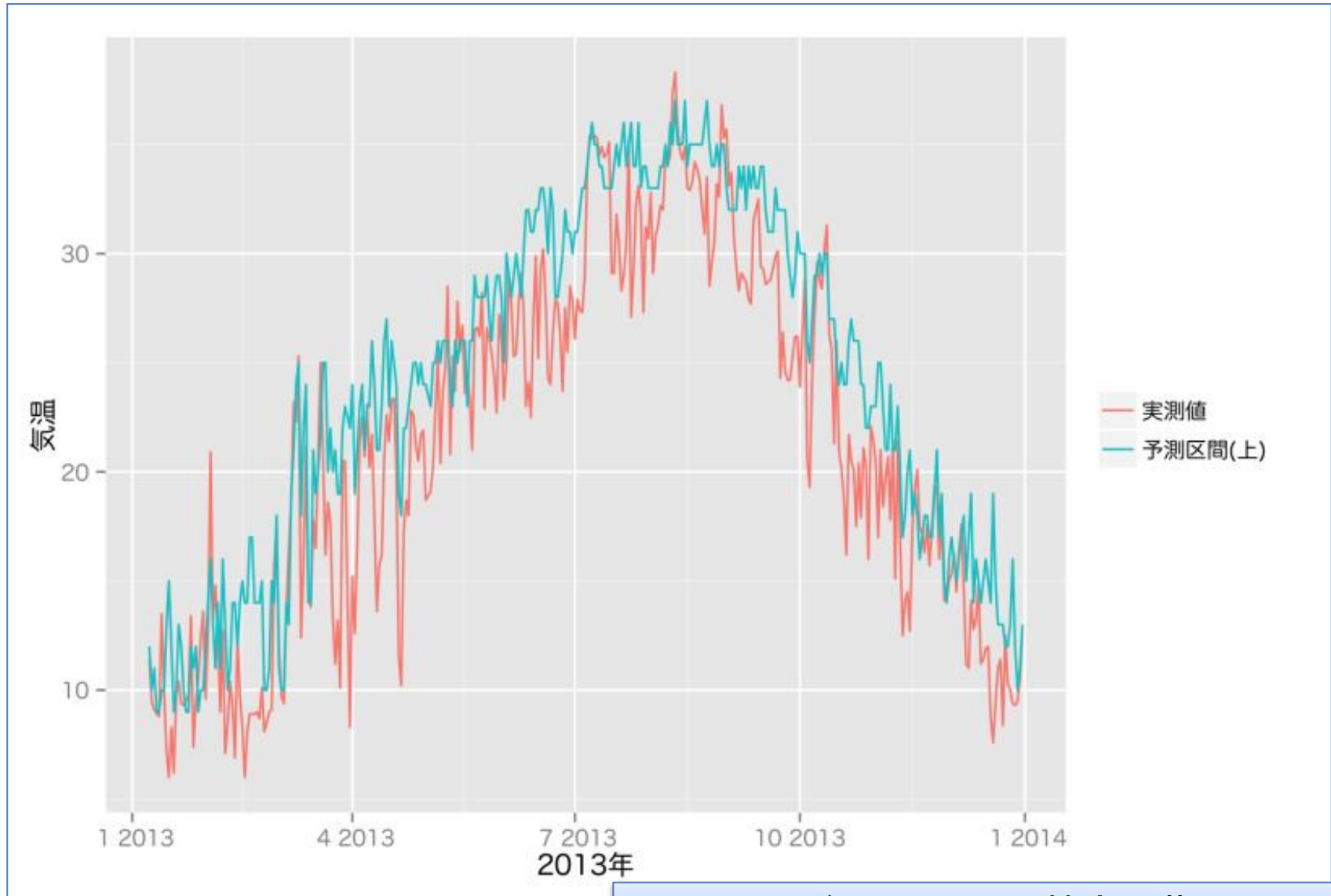


7日後の予測最高気温と実測値



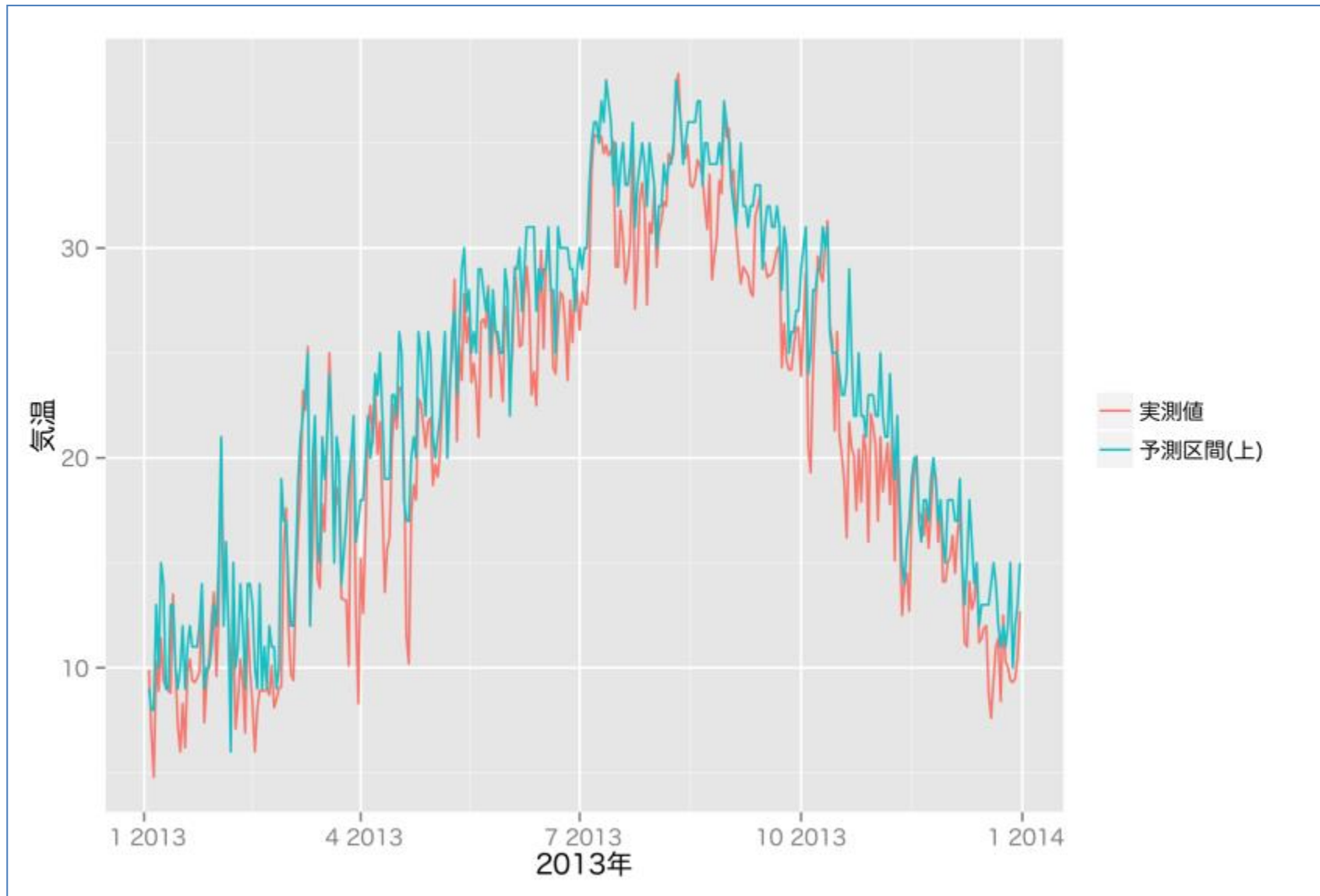
2日後に比べて、予測範囲の幅が大きい

7日後の予測最高気温と実測値

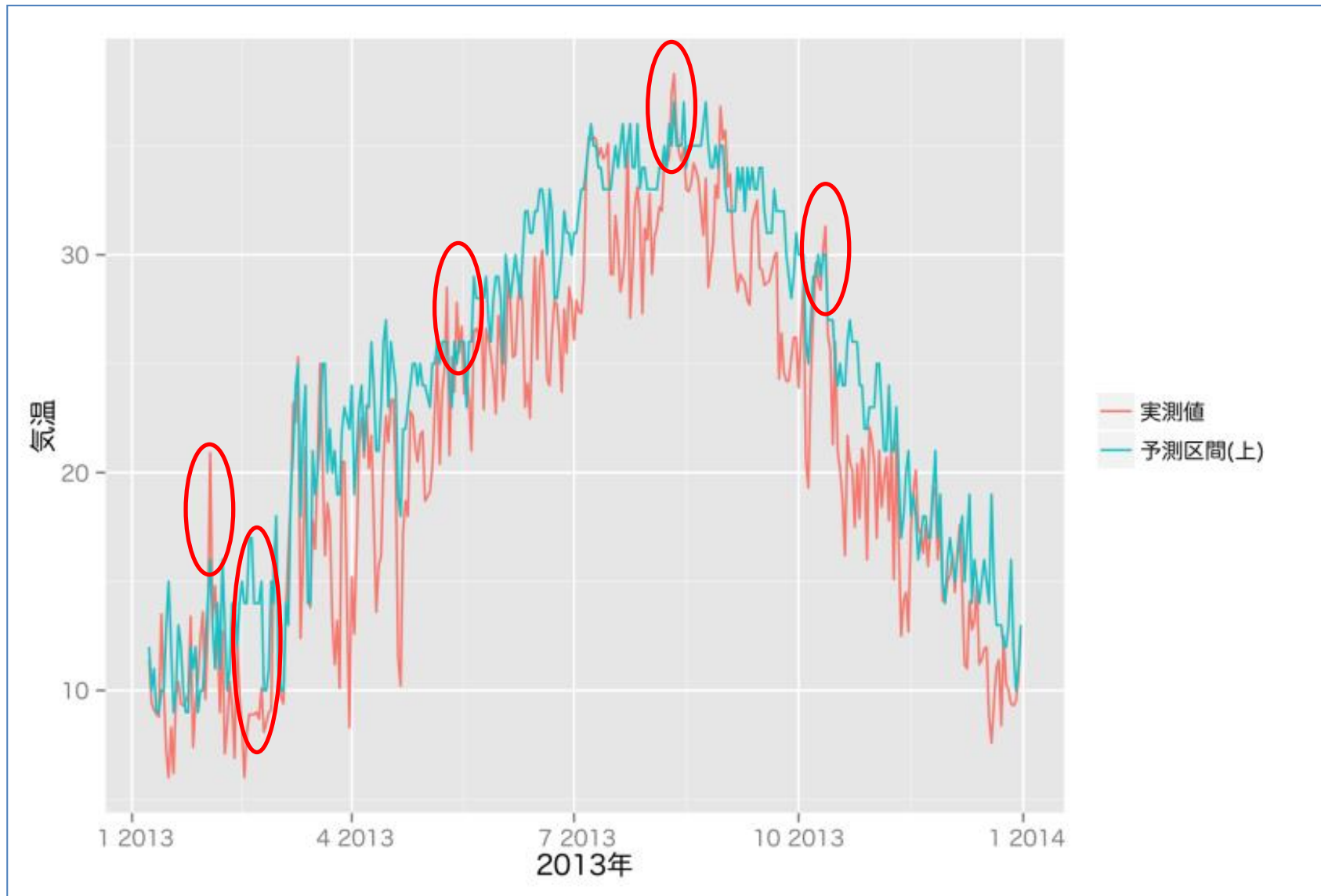


2日後に比べると、精度は落ちる

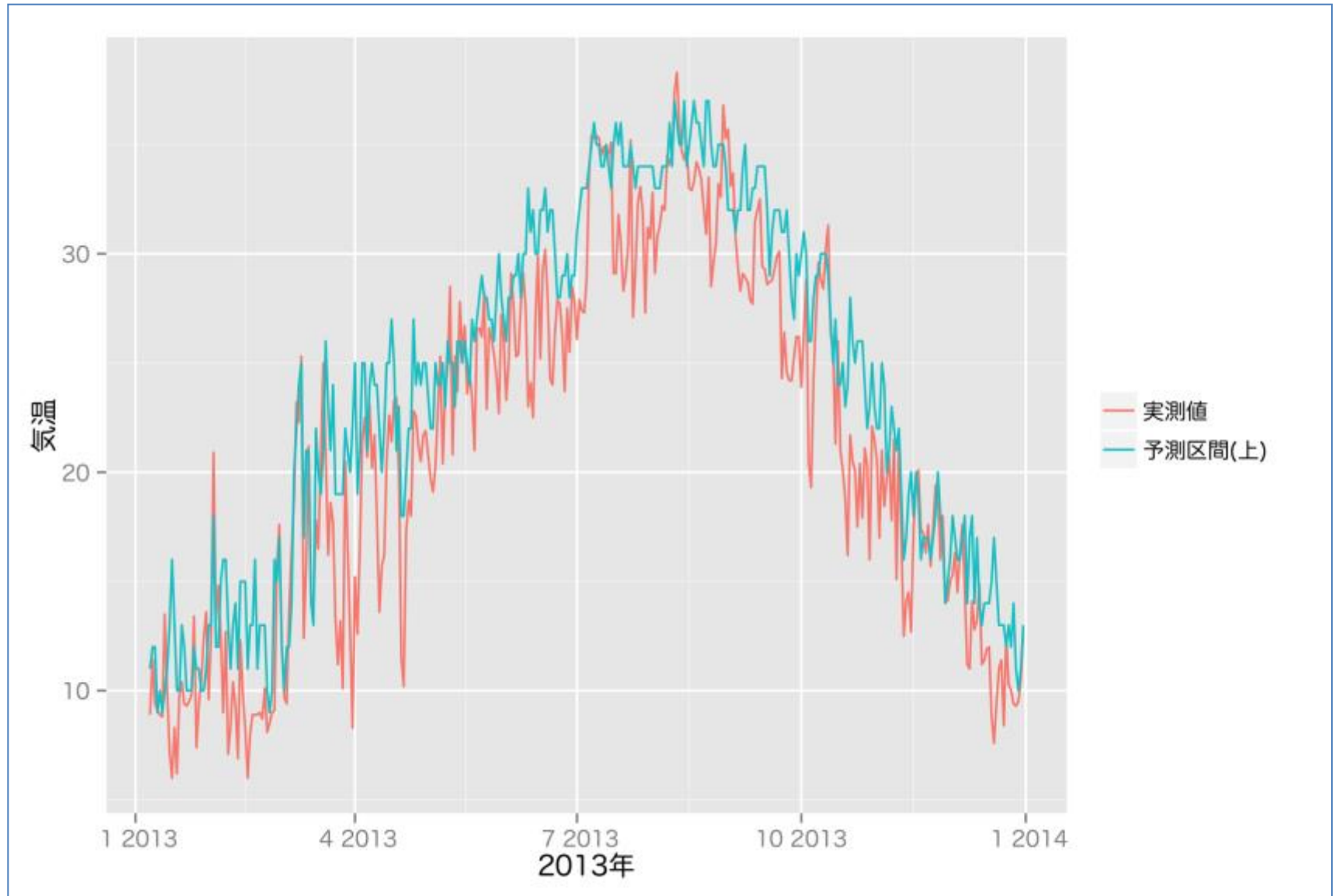
2日後の予測最高気温と実測値（再掲）



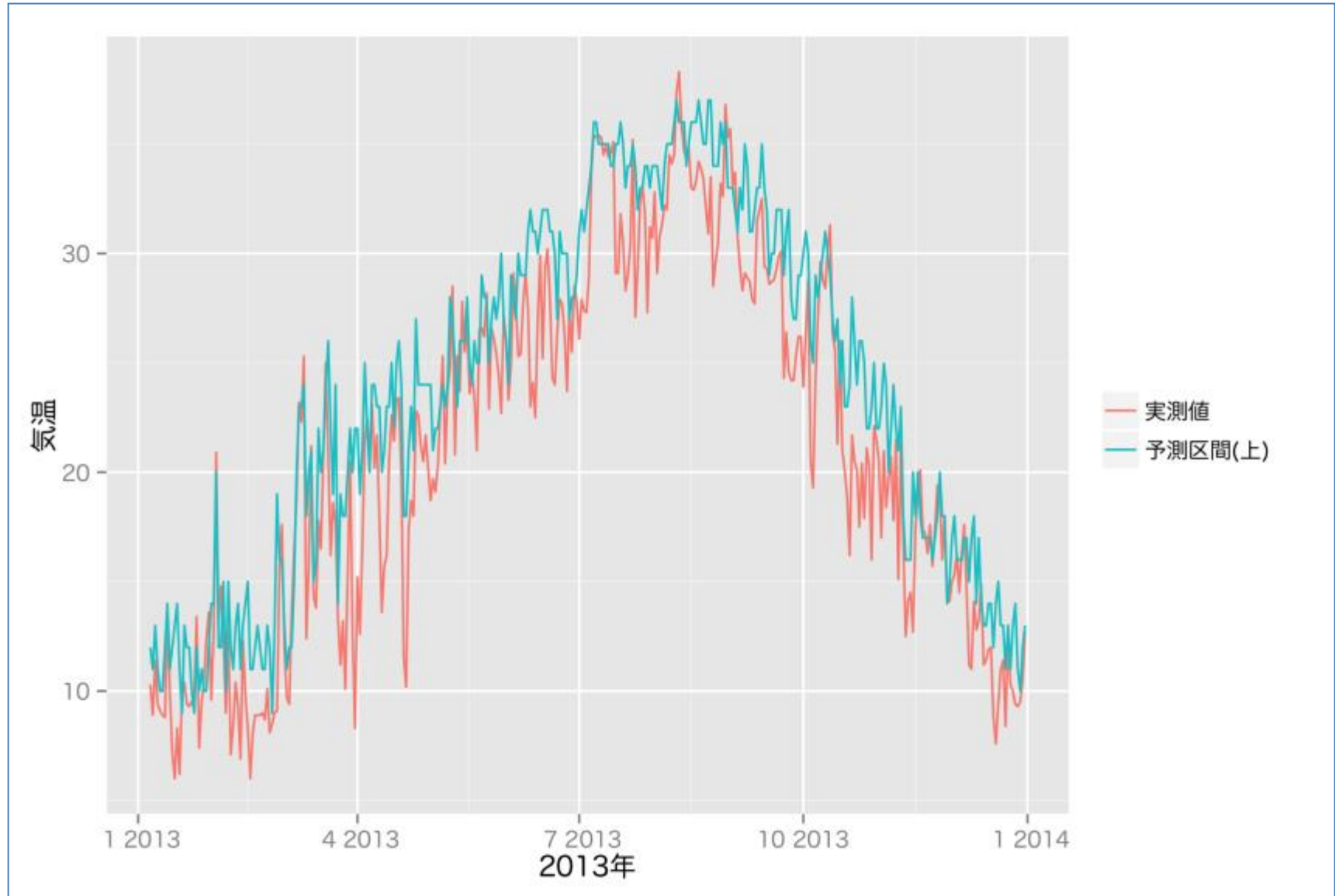
7日後の予測最高気温と実測値



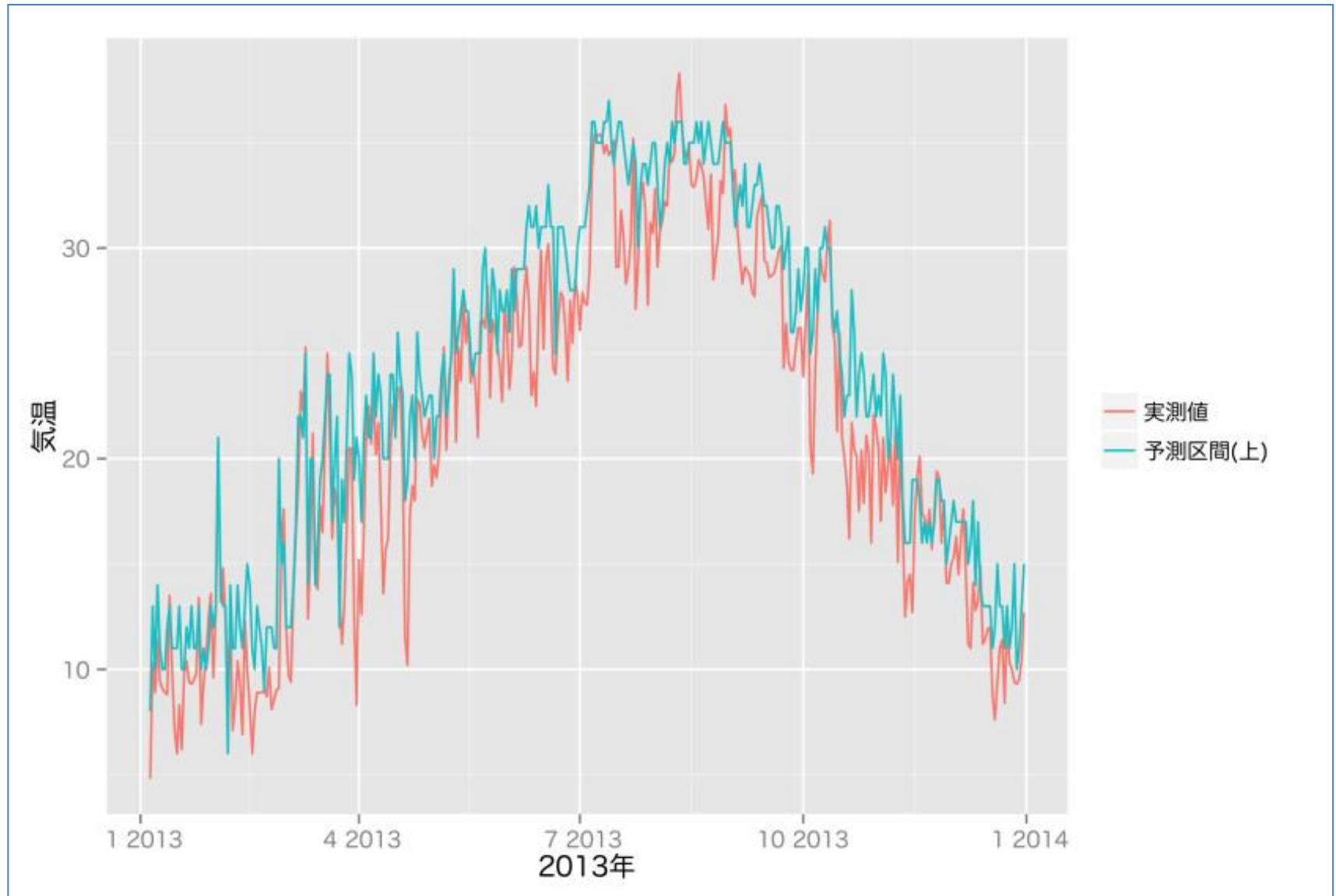
6日後の予測最高気温と実測値



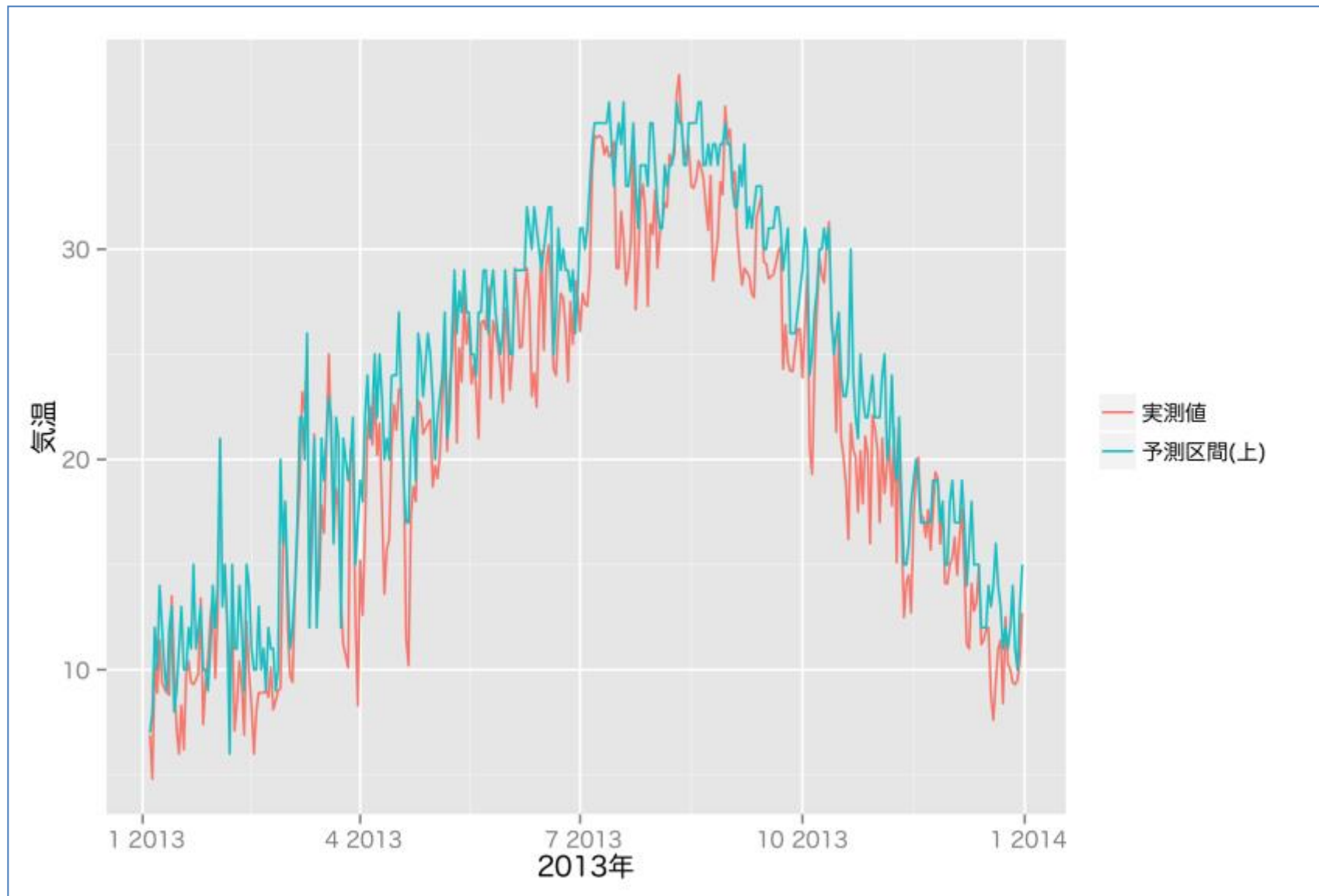
5日後の予測最高気温と実測値



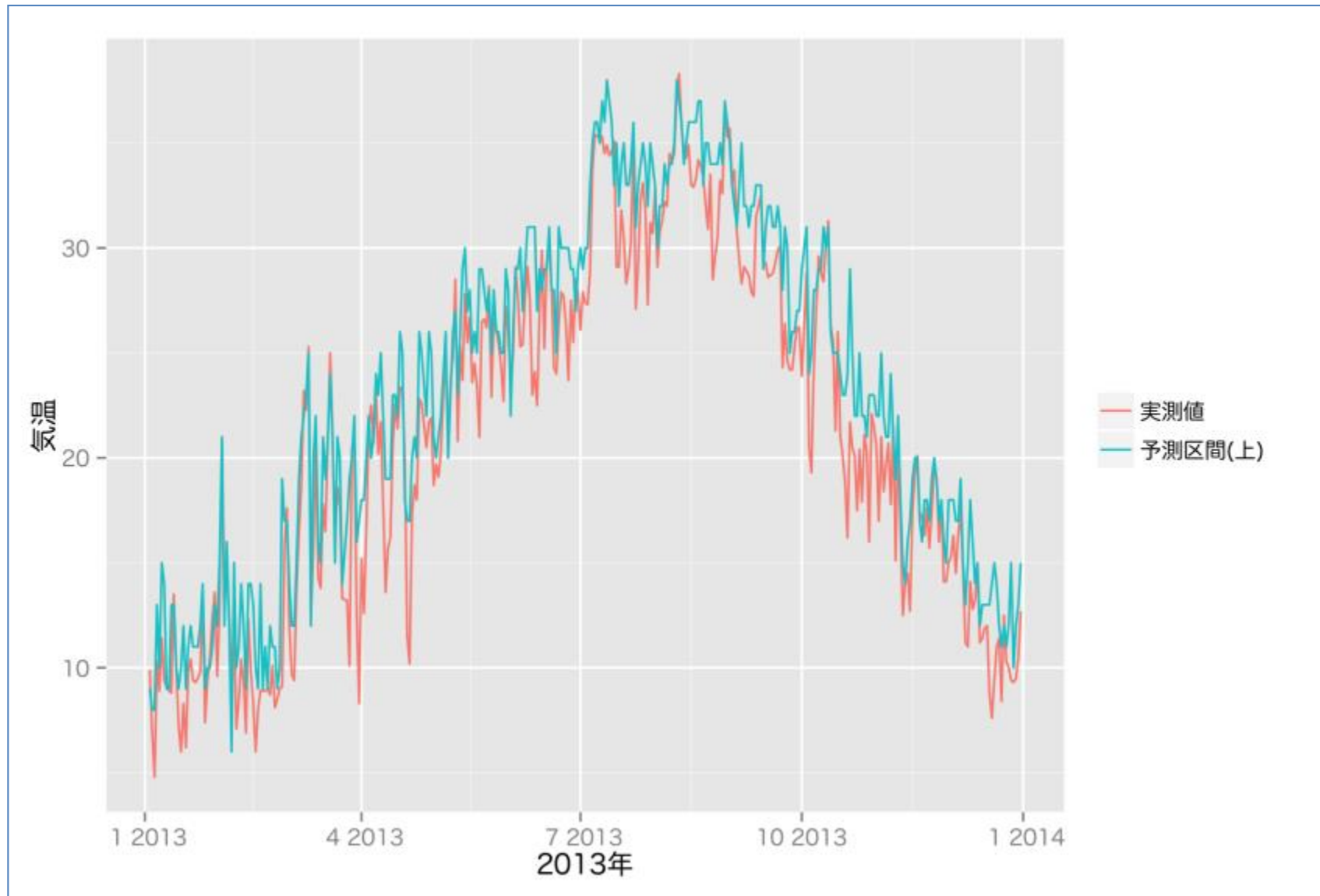
4日後の予測最高気温と実測値



3日後の予測最高気温と実測値



2日後の予測最高気温と実測値



	上限超過日数	上限超過日数(1℃以上)	最大超過(℃)
2日後	52日	17日	3℃
3日後	49日	18日	3℃
4日後	57日	22日	3℃
5日後	60日	25日	2.8℃
6日後	67日	31日	4℃
7日後	62日	30日	4.9℃

- 上限を超えた日数のうち、多くが1℃未満の超過であった
 - 上限は1℃程度のマージンを見た方が良い
- 6日後、7日後の予測データはやや精度が落ちる
- 4日後～2日後は安定して高い精度であった

- 気象庁XMLによって天候と相関のあるデータについての需要予測が可能となる
 - 今回は電力使用量について、需要予測を行なって来たが、他のデータにも応用することが出来る
- 気象庁XMLの他にも公開されているデータが増えて来ており、組み合わせることによってより精度の向上が期待出来る
- 気象庁XMLでは他にも多くのデータが含まれており、データの特性に合わせた利用方法が考えられる
 - 今回は週間天気予報のデータのみ利用
- 気象庁XMLによるデータ配信と共に、気象統計情報の公開によって、気象データが身近なものとなり、データ利用の活性化が期待出来る

ご静聴ありがとうございました