

2012年度卒業研究

那須風速データの混合分布モデルによる解析

卒業研究「那須気象データの統計モデルによる体系化」より

宮加谷 祥子

大森 興太郎・酒井 優行・高山 大樹・高橋 礼穂・矢野 真知子との共同研究

指導教員：遠藤 敏喜・島津 秀康

概要 本研究では1942年から1998年までの那須地域における風速データを解析した。データは自由学園那須農場の農場員によって57年間に亘って観測されたものである。風速データでは無風観測日数が1980年代から著しく増加していたが、この事実と風特有の季節性を取り込んだ混合分布モデルを構築した。モデルが捉えた結果をふまえて、風の減少と風速低下に機器の経年劣化や観測環境の変化が関わっている、那須地域における風の季節変動には局地的な季節風の影響があると結論付けた。

1 はじめに

1.1 卒業研究について

2012年度テーマ別グループ研究“数理モデルとインターフェイス”の数理チームの卒業研究[5]では、自由学園那須農場の山口泉氏らが観測して残した気象資料をデータ化し、統計解析を行った。気象観測資料は1942年から1998年の57年間20,819日分存在し、欠測日はほぼ皆無である。データ化にあたり最高気温、最低気温、地中温度、地下水位、雲量、雲形、最大風向、最大風速、定時風向、定時風速、降水量、視程、天気、湿度ほかの計29の観測項目を抽出したが、手帳にはその他の記録も含まれている。気象観測を戦前から那須地域で行っていたのは他に畜産草地試験場だけであり、地域の記録として社会的にも意義のある活動であったと言える。メンバーの6人は、それぞれの関心から最低気温、最高気温、地下水位、降水量、視程、風向、風速について統計解析を行った。これらの項目に関しては、関係者へのインタビュー、観測方法の確認、他の観測地で記録された数値との比較などを行うことでデータの信頼性を確認した。本稿では、紙面の都合上、卒業研究[5]の中から定時風向と定時風速に関する結果のみを紹介する。他の研究については[2, 6, 9, 15, 16]を参照されたい。

1.2 風に関する先行研究

風データの解析は台風シミュレーション[4]や風力発電設置の基礎調査[7]などとして行われているが、ここで示される風速の分布は日本全体規模の調査結果であることが多く、地域を限定して風速を調査したものは少ない。那須地域に関しても、風向の研究[8, 17]は存在するが、風速に関するものはない。本研究では那須地域の風の季節変動を丁寧に捉えることにより、局

地的な風の特徴を詳細に明らかにする風速モデルを構築する。

風は気圧の高い場所から低い場所へと吹き込み、季節に応じて気圧は変化している。よって風の変動に季節性があることを想定する。実際、日本において地上付近の風は年間を通して季節風に支配されている。また地上から約1kmまでは大気境界層と呼ばれ、地表面摩擦や地形の影響を強く受けるために局地性が強くなる傾向にある[7]。冬は大陸が高気圧、海洋が低気圧となるため北海道から九州では大陸から海洋に向かって北西の季節風が吹き、夏は逆に太平洋高気圧が発達するため大陸が低気圧となり、太平洋から大陸に向かって南から南東にかけての風が吹く。春と秋は移動性高気圧、低気圧が日本を通過し、それに伴って風向風速は変化する。風速の季節変動は北海道から九州までの範囲は冬に風が強く夏に弱いとされ、冬は東北地方が最も強く、夏は北に進むにつれて風が弱くなる。那須地域においては、冬季に“那須おろし”と呼ばれる北または北西の強い局地風が発達するため[8, 17]、冬は風速が強い傾向となることが予想される。

1.3 モデルについて

本研究では、まず局所回帰平滑化を用いたモデルによって、次にその改良である混合分布モデルによって風速の季節変動を明らかにするとともに、モデルが捉えた那須地域における風の傾向を考察する。

局所回帰平滑化はデータ x の近傍で局所的に多項式を最小2乗法で当てはめ、不規則な変動などを取り除き、データの持つ傾向や規則的な変動を明らかにする。すなわち、ある範囲の重みをつけたデータを最小2乗法によって平滑化した値を与える。風速の分布は人間やモノの寿命の分布としてよく用いられるワイブル分布によって評価できることが一般的に知られている[4]が、本研究ではワイブル分布と同じ族である指数分布を採用する。指数分布はワイブル分布よりもパラメー

タが1つ少ない単純な確率分布であるため、よりシンプルなモデルを構築することが可能となるからである。混合分布モデルとは2つの分布を混ぜた分布をいう [1]。データは様々な情報を持っており、必ずしもひとつの分布でその情報を記述しきれるとは限らないためそれぞれ異なる意味の分布を考慮し記述、混合することで、よりデータの傾向を適切に反映したモデルを構築することができる。以上の詳細については [11] を参照されたい。

1.4 風速データ概要と無風観測日数の増加

調査対象である風速データは、風杯型風速計を用いて1日1回、毎朝9時時点の風速を記録していたもので、57年間を通しての最小値0、第1/4位0.2、平均値1.4、第3/4位2.2、最大値13.2(m/s)であった。

図1は観測に使われていた風速計と風向計の写真である(1953年4月に山口泉氏撮影、山口曜氏提供)。高さ約10mの鉄塔に設置された風速計は地上にある記録装置と連動しており、観測された風速は自動で記録されていた。観測者は自動記録された数値を読み取り、手帳に記録していた。



図1: 風速/風向計

風速データに見られる特徴として、図2のように1984年頃から風速0である無風の観測日数＝ゼロデータが顕著に増加している。気象観測が始まった1942年の無風日数は25日だが、1998年には年間で307日も無風が記録されている。大型の台風襲来頻度の減少などにより、1960年以後から日本において風速は低下の傾向にあると言われている [12] が、無風観測日数の増加傾向が必ずしも那須地域において無風状態が増えた結果と言えるわけではない。仮に那須地域において風が

減少していたとしても、自然現象として無風状態の増加がこれほど急激な傾向をとることは考えられにくい。実際、気象観測を手伝ったことのある方から、“風速計周辺の樹木が成長し、風を防ぐ障壁となった。もしくは風速計の経年劣化による風速観測への障害といった要因が背景にあるのではないか”という見解も伺った。つまり、自然現象以外の要因がゼロデータの増加に影響している可能性が高いと言える。

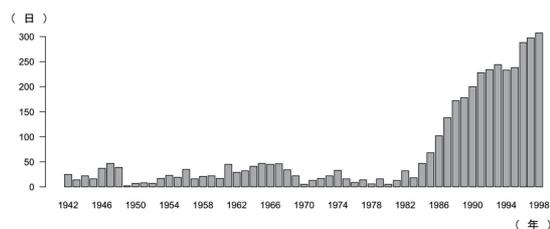


図2: 無風観測日数の変化

栃木県では他の気象観測所でも風速観測が行われているため、那須農場と同様の傾向が見られるか確認をした。気象庁のホームページ (<http://www.jma.go.jp/>) より、比較的那須農場に近い大田原、黒磯、塩谷、真岡の4ヶ所の観測所を取り上げ、1976年、1980年、1986年、1998年の各年の風速データから無風観測日数の変化を調べて比較した(図3)。すると那須農場と同様、無風観測が増加している傾向が見られた。

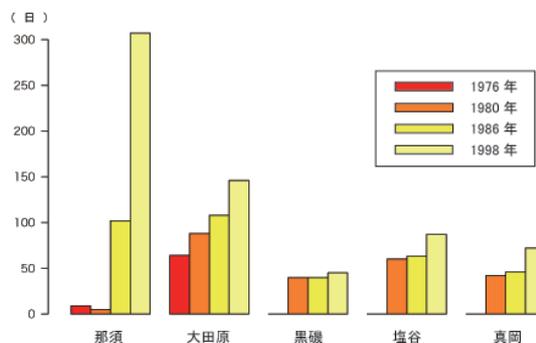


図3: 他観測地点の無風観測日数変化

2 ゼロデータを考慮しない風速モデル

風速の経年変化を調べるために、風速 y_t を、長期的な緩やかな変動 S_t と、月効果 $M_{k(t)}$ の2成分に分解して $y_t = S_t + M_{k(t)} + \varepsilon_t$ とした [10]。ここで S は局所回帰 [3] によって定められた平滑化関数で平滑化幅は1年とした。 $M_{k(t)}$ は観測時点 t ごとに $k = 1, 2, 3, \dots, 12$ までの異なる月を水準とする定数であり、この定数を足すことにより期待値に月ごとの変動すなわち季節変動を与える。また、 ε_t は残差を示す。

このように風速データに局所回帰平滑化をかけ、データの分解を行った。長期的で緩やかなトレンドはデータ全体の傾向を抽出することはもちろん、その後データがもつ周期的な傾向に分解する際、傾向を拾いやすくすることも目的としている。

分解した各トレンドを図4に示す。長期トレンドは風速を平滑化した値、短期トレンドは長期トレンドの平滑化値上の値の振幅を示している。各トレンドの線の形状に注目すると、長期トレンドでは徐々に低下していく傾向、そして短期トレンドには周期的な波形が見られ、これらからそれぞれゼロデータの増加に伴う風速の低下傾向と、50年分の季節変動を確認した。最下段の図はモデルから算出された風速期待値の係数を月ごとにプロットしたものである。つまり年間の季節変動を表しているのだが、春季に上がり夏季に下がるという季節変動が示されているため、一見風速データの持つ傾向を表すことができているように見える。

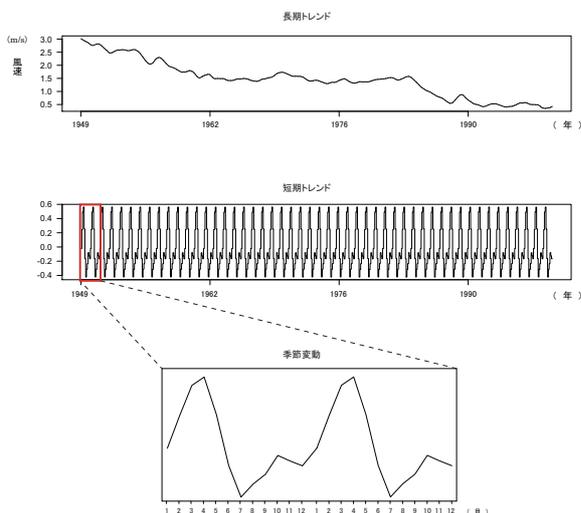


図4: 長期、短期トレンドと季節変動

しかし2つのトレンドを抽出した残差を年ごとにまとめた箱ひげ図から検証すると、1990年頃からはずれ値が増える傾向が残っており、残差の分散が経年で変化していることが分かる。これは平滑化の際、トレンドがゼロデータに引きずられ、結果、外れ値が多くなったと考えられる。これらのことから、単に平均的傾向に関するモデルだけでなく、分散も考慮したモデルの必要性が示唆される。

3 混合分布モデル

3.1 モデル

本節では風速ゼロの発生確率を区別して解析を行い、風速の分布を記述する混合分布モデルの構築について

述べる。時点 t での風速 y_t をゼロと非ゼロの場合に分け、ゼロの場合は風速のモデルをゼロの発生確率 P_t で表し、非ゼロの場合は風速が指数分布に従うと仮定する混合モデルで記述する。すなわち風速 y_t の従う確率密度関数 $f(y_t; P_t, \lambda_t)$ を次式で定める。

$$f(y_t; P_t, \lambda_t) = \begin{cases} P_t & (y_t = 0) \\ (1 - P_t)\lambda_t e^{-\lambda_t y_t} & (y_t > 0) \end{cases}$$

ここで、 P_t, λ_t はパラメータである。今回はこのパラメータ P_t, λ_t が2章のモデルで示したものと同様に、長期的変動 S_t と月変動 $M_{k(t)}$ であり、 $\logit(P_t) = S_t, \lambda_t = S_t + M_{k(t)}$ のようにモデル化する。ここで $\logit(P_t)$ はゼロデータの発生確率を表す P_t のロジット関数である。また、風速パラメータ λ_t の S と M_k は2章のモデルで示したものと同様に平滑化関数と月効果を指す。

当初長期トレンドの平滑化関数は、短期トレンドの月効果に対して年の効果に設定していたが、モデル当てはめ後の残差の分布がよくなかったため、細かい変動も拾えるよう時系列で平滑化をかける平滑化関数に変更した。しかし平滑化幅を細かく設定するほど平滑化を施すデータの範囲が狭まり、第1項の平滑化で細かい変動まで分解されてしまうため、季節変動といった周期的な波形が見えづらくなってしまった。そこで周期的な波形が視覚的に確認できる範囲で、最小の平滑化幅となる平滑化関数に設定することとした。

この混合分布モデルの期待値 $E[Y_t]$ と分散 $\text{Var}[Y_t]$ は以下の通りである。通常、指数分布の期待値は $1/\lambda$ 、分散は $1/\lambda^2$ となるが、本モデルではゼロデータの混合モデルであるため、風速0の発生確率 P_t で割り引かれたものとなる。

$$E[Y_t] = (1 - P_t) \frac{1}{\lambda_t}, \quad \text{Var}[Y_t] = \frac{1 - P_t^2}{\lambda_t^2}$$

混合モデルにすることで、全体的な傾向であるゼロデータ増加の影響と、短期トレンドである季節変動を同時に見ることができる。また、モデルはゼロの発生確率と風速の分布を混合してあるため、風速のみの分布だけでなく、風速データとゼロデータの相互関係を考慮した風速の分布といった結果も必要に応じて自由に取り出すことが可能となっている。

3.2 パラメータの推定

パラメータ P_t と λ_t は最尤推定法により推定される。混合分布モデルの対数尤度は、 n をゼロの個数、 m を非ゼロの個数として、 $\ell = \log L(P_t, \lambda_t; y_t) = n \log(P_t) + m \{ \log(1 - P_t) + \log(\lambda_t) \} - \lambda_t \sum_{y_t > 0} y_t$ で表される。この式を整理し、 P_t, λ_t のそれぞれについて偏微分をす

ると尤度方程式が求まる。ここで、 \bar{y}_t は非ゼロデータの平均を表す。つまり P_t と λ_t はそれぞれ2項分布と指数分布の最尤推定量となる。また、 λ_t は標本平均の逆数でもある。このようにパラメータ P_t と λ_t の最尤推定量は無関係であり、片方のパラメータがもう片方のパラメータに影響を与えない。よってゼロと非ゼロの場合分けによって別々にパラメータの推定ができる。

$$\frac{\partial \ell}{\partial P_t} = \frac{n}{P_t} - \frac{m}{1 - P_t} = 0 \rightarrow P_t = \frac{n}{n + m}$$

$$\frac{\partial \ell}{\partial \lambda_t} = \frac{m}{\lambda_t} - \sum_{y_t > 0} y_t = 0 \rightarrow \lambda_t = \frac{m}{\sum_{y_t > 0} y_t} = \frac{1}{\bar{y}_t}$$

3.3 風速ゼロの発生確率

図5はモデルによって明らかとなったゼロの発生確率を示し、風速ゼロの発生確率 P_t が時間の経過と共に上昇していく様子が分かる。1976年までは緩やかな傾向となっており、その後大きく上がっている。これは図2の無風観測日数の増加傾向とかなり近い形状である。最後再び緩やかな傾向となったのは、無風の日数が300日を越え、無風日数増加の傾向が伸びにくくなったからであると考えられる。

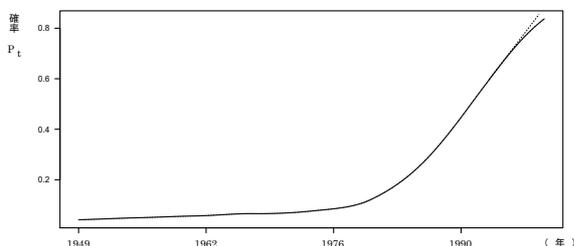


図5: ゼロデータの発生確率

3.4 風速の期待値

図6はゼロデータを除いた風速の月ごとの期待値を示したものであるが、全体の傾向として無風状態を除いても風速は年の経過と共に低下していることが分かった。グラフが途切れている箇所は風速が0m/sの日を示し、1990年より少し前からゼロデータがかなり増えている様子がこの図からも分かる。この風速の期待値を月ごとにまとめ、箱ひげ図で表したものを図7に示す。4月にピークを迎え、7月最も風速が低くなるという季節変動は4.3節で紹介したモデルの結果と同じ傾向となった。[7]より、風速は夏季に低く冬季に高い傾向にあると言われているが、これと同じ傾向は本研究

でも確認され、さらに細かくは那須地域においては春季に最も高くなるという結果も読み取れた。

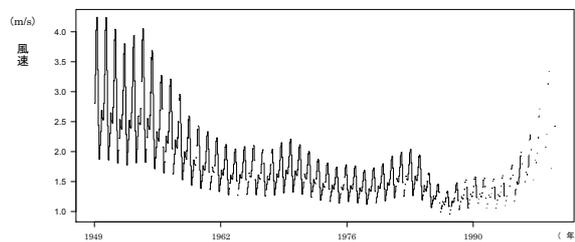


図6: ゼロデータの影響を除いた風速期待値

混合分布モデルによって定められた風速の期待値と分散の時系列変化を眺めると、先に示したゼロデータの影響を受けていない風速の期待値と同様、期待値と分散ともに季節変動を示す短期トレンドの波形が見られる。終盤にかけて期待値の波形の振幅が小さくなり、これは混合モデルによりゼロの発生確率が高くなるほど割り引かれる値が大きくなるからである。1980年代後半からゼロデータの数が顕著に増加することによって風速のデータ数は減少する。365日中、約300日がゼロデータという年もあるが、期待値の波形は小さくなっているものの、季節変動を示す。

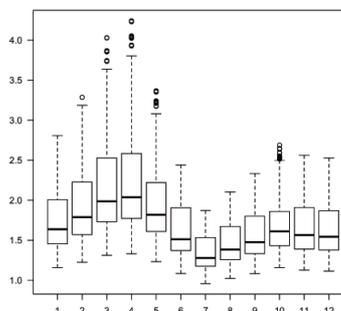


図7: 風速の年間変動

3.5 考察

図6の風速の期待値が示すように、データから無風状態を取り除いても風速は年の経過と共に低下傾向にあることが分かった。年の経過につれて大きく期待値が低下しているのは樹木などの障壁によって風が防がれたことや、劣化によって風速計がうまく回らなくなり、風速が低く観測されるようになったのではないかと考えられる。

1.4節で紹介した那須農場以外の4ヶ所の観測地点でもゼロデータを除いた風速の分布を年ごとにみると、

表 8: 季節ごとの風配図

季節	春	夏	秋	冬
月	3・4・5月	6・7・8月	9・10・11月	12・1・2月
風配図				
データ数	4,473	4,112	4,041	4,172
卓越風と割合	北 10.3% 南 20.3%	南 25.3%	北 13.7% 北東 11.7% 北西 10%	北 16.1% 北西 15.1% 西 12.6%

データ数における風速 0～2m/s の頻度の割合は全ての場所で一貫して高くなっている。また 1980 年は風速 6m/s 以上の比較的強い風が何度か観測されているが、1998 年にはほとんど観測されていない。これらの観測所は無人観測所であるため、観測場所の周辺環境や機器の整備が定期的に行われている可能性が低いことから那須農場と同様の傾向が見られたのではないかとと思われる。

また、期待値と分散は変動しているが、1990 年以後、期待値は低下、分散は上昇というそれまでの傾向とは異なる変化をしている。これは 1990 年中頃から無風観測日数の増加傾向の伸びが若干緩やかになったことで、年間の風速データ数が増えた。しかし風速のデータ数は増えても風速は低下の傾向にある、つまり風速は低下しながらもデータの分散は増えていることになる。よって風速の期待値は低下、分散は上昇の傾向となったと考えられる。

4 風向について

3.4 節で述べたとおり、那須農場の観測データにおける風速の季節変動の傾向は [7] とは少々異なる傾向であった。その原因に那須地域における局地的な要因があるのではないかと考え、地理的な影響が現れる風向データのもつ季節性を検証した。

栃木県では、夏季は南または南東の風、冬季は北西または北の風が卓越すると言われている [13]。[8, 17] では、那須野ヶ原扇状地における屋敷森 (宅地防風林) の分布を指標として局地的な卓越風を推定する調査が行われている。那須野ヶ原扇状地は冬季、那須連山から吹き下ろす“那須おろし”という強い局地風が発達している。屋敷森とは家屋や耕地を強風や雪から守る目的で設置される林で、那須農場が位置する扇状地の西部は北または北西に屋敷森を構えているため、冬季は北または北西の強風が吹いていることが分かる。

同農場では風速と共に風向の観測も行われており、観測野帳に記録が残されている。今回風速データの検証結果を風向データと照合し、那須地域における風の傾向を考察した。風向の観測方法は [14] を参考に、矢羽根型風向計が示す 16 方位と、風速が 0.2m/s 以下を示す“静穏”を含めた 17 種類で記述される。

表 8 はデータ数における各方位が占める割合を風向の頻度とし、1942 年から 1998 年まで、20,189 日分の風向データを風配図を用いて季節ごとの卓越風を明らかにしたものである。

卓越風はそれぞれ夏季は南、冬季は北、北西、西となり、[13] で述べられている季節風や局地風の影響を受けた傾向が那須農場のデータにも見られた。また春季は夏季、秋季は冬季と似た風向の分布となった。

風速と風向で顕著な特徴の現れる時季が異なるが、それらを照合すると、北よりの風は比較的風速が高く、南よりの風は風速が低めであるということが分かった。本モデルにより、那須地域において最も風速が高くなるのは 4 月、低くなるのは 7 月という季節変動が明らかとなった。表 8 で示した風配図では、春季と夏季の風向の分布は似た形となっている。しかし春季にも那須おろしなど比較的強い北よりの風が度々吹くことで風速が高い傾向となり、また夏季は南の風が非常に卓越することから風速が低い傾向となるのではないかと考察した。

5 結論

本研究は那須農場で観測された風速・風向データを用いて那須地域における風の傾向を解析し、風速データが有する特徴を考慮した風速の混合分布モデルを構築した。モデルを通して風の季節変動、風速の低下や年の経過と共に増加するゼロデータの傾向を確認した。まとめると以下のようなになる。

風の傾向については、(1)年の経過と共に風速0m/sの発生確率は高まり、風速の分布も弱い風が卓越するようになる。他の観測地点と比較しても無人観測所で同様の傾向が見られたため、機器の劣化や樹木の障害といった観測環境の不備が影響していると考えられる。(2)北よりの風は比較的風速が高く、南よりの風は風速が低めの傾向にあり、南の風が卓越する夏季は風速が低く、春季は比較的強い北よりの風が度々吹くことから風速が高い傾向となる。つまり那須地域においてピークを迎える春季(4月)は、北よりの風が卓越しており、那須野が原に吹く強い局地風の影響であると言える。

モデルについては、(1)風速のみを示すモデルでは風速を捉えきることができなかつたのはゼロデータの影響である。ゼロデータをあらかじめ引いておくことによって風速の分布を捉えられるモデルとなった。(2)混合分布モデルへの当てはめにより、ゼロデータの発生確率が急激に増加する様子やそれに伴う風速の期待値の低下が見られた。これはそれらの傾向の要因が風速計の劣化や、樹木の成長が風観測の障害となっていることを提起している。

近年那須地域において風が吹かなくなっている、または風が弱まっていると思われるような傾向も見られたが、本研究では気象観測にかかわった方の経験や知識に基づき、樹木が風の障壁となっていることや観測機器の劣化が、ゼロデータ増加と風速低下の大きな要因であると解釈することとした。混合分布モデルにより、“ゼロデータの増加”というデータの特徴が経験や知識と一致するような結果を示すことができた。

参考文献

- [1] Geoffrey McLachlan and David Peel, *Finite Mixture Models*, Wiley-Interscience, 2000.
- [2] Machiko Yano, Hideyasu Simadzu, and Toshiki Endo, Quantifying the effect of heat on milk production by individual cows of different calving conditions, submitted (May 2013).
- [3] William S. Cleveland and Susan J. Devlin, Locally weighted regression: An approach to regression analysis by local fitting, *Journal of the American Statistical Association* 83 (1988), 596–610.
- [4] 天野貴文, 台風シミュレーション・システム開発と風向別最大風速の評価, *Theory and Applications of GIS 14* (2006), 145–155.
- [5] 大森興太郎・酒井優行・高橋礼穂・高山大樹・宮加谷祥子・矢野真知子, 那須気象データの統計モデルによる体系化, 2012年度自由学園最高学部卒業研究, 2013.
- [6] 大森興太郎・酒井優行・高橋礼穂・高山大樹・島津秀康・遠藤敏喜, 栃木県那須地域の気象データを用いた統計モデルとデータ解析 その2, 第9回数学総合若手研究集会テクニカルレポート集 (2013), 285–288.
- [7] 河野仁, 日本列島の風速分布および地形と風速・乱流の関係, *環境技術* 41 (2012), 544–549.
- [8] 小園修, 那須野ヶ原北西部における「那須おろし」の風系分布, *東北地理* 35 (1983), 20–25.
- [9] 酒井優行・高山大樹・島津秀康・遠藤敏喜, 栃木県那須塩原市蛇尾川の出水と地下水位変動モデルについて, *応用数学合同研究集会報告集* (2012), 236–237.
- [10] 島津秀康・柴田里程, 局所回帰による時系列の分解から明らかになった野鳥羽数の環境要因変化との関連性, *日本統計学会誌* 34 (2005), 187–207.
- [11] J. M. チェンバース・T. J. ヘイステイ編集, 柴田里程訳, *Sと統計モデル—データ科学の新しい波*, 共立出版, 1994.
- [12] 鄭永培, 風観測および実験による風速のモデルに基づく設計用風荷重評価, 学位論文(東京大学), 1997.
- [13] 栃木県生活環境部環境局環境政策課, 栃木県環境基本計画(改訂版), 栃木県, 2006.
- [14] 三浦栄五郎, 気象観測法講話, 地人書館, 1944.
- [15] 宮加谷祥子・矢野真知子・島津秀康・遠藤敏喜, 栃木県那須地域の気象データを用いた統計モデルとデータ解析, 第9回数学総合若手研究集会テクニカルレポート集 (2013), 155–158.
- [16] 矢野真知子・島津秀康・遠藤敏喜, 個体差と気象条件を考慮した搾乳量と乳成分率のモデリング, *形の科学会誌* 27 (2012), 148–149.
- [17] 吉野みどり・井出信子・大平菜穂, 那須扇状地における屋敷森の分布と風向, *文教大学教育学部紀要* 15 (1981), 1–8.