

情報理論を用いた天気予報の必要性の定量化

理学専攻・情報科学コース 2240666 服部佳奈

1 はじめに

気象情報とビジネスは切っても切り離すことができない関係であり、気象によって収益が左右されることがある。天気予報は経済的に有用であり、高精度な気象予報をすることは非常に重要である。そこで、予報技術や結果の検証を行なっていく必要がある。先行研究 [1][2] では、情報理論におけるエントロピーの考え方を導入することで気象予報技術の精度を検証する尺度を定式化して、簡潔で理解しやすい検証方法が提言されている。そこで本研究は、天気予報の必要性が高い地域・季節を情報理論を用いて特定し、天気予報の効果の検証に役立てることを目的とする。

2 データと計算方法

2.1 データ

- ・気象庁 日別降水量データ 1976-2022 年
- ・気象庁 日別天気概況データ 1997-2022 年
- ・GPCP Daily Precipitation 1997-2022 年

2.2 計算方法

情報理論においてエントロピーは次のように定義される。

$$\bar{I} = - \sum_i (p_i \log_2 p_i)$$

次に、降水の有無の確率を、日別データを用いて降水のあった日数をその月の日数で除することで定義する。これを用いて計算したエントロピーを、

$$\bar{I} = -p_s \log_2 p_s - p_r \log_2 p_r$$

p_s : その月に降水がなかった確率,

p_r : その月に降水があった確率

とし、本研究では天気予報エントロピーと呼ぶ。

天気概況データから“晴”という文字列を含む場合は〈晴れ〉、“雨”という文字列を含む場合は〈雨〉、“曇”という文字列を含む場合は〈曇〉と分類し、以下のよう

$$\bar{I} = -p_s \log_2 p_s - p_r \log_2 p_r - p_c \log_2 p_c$$

p_s : その月に〈晴れ〉が出現した確率

p_r : その月に〈雨〉が出現した確率

p_c : その月に〈曇〉が出現した確率

天気カテゴリーを9つの分類で分ける場合には、各分類がその月に出現した確率 $p_n (n = 1, 2, \dots, 9)$ を用いて

$$\bar{I} = - \sum_{i=1}^9 (p_i \log_2 p_i)$$

	分類	定義
1	〈時々晴〉	「時々晴」という文字列を含む
2	〈一時晴〉	「一時晴」という文字列を含む
3	〈晴〉	「晴」「降水なし」という文字列と一致する
4	〈時々雨〉	「時々雨」「時々雪」「時々みぞれ」という文字列を含む
5	〈一時雨〉	「一時雨」「一時雪」「一時みぞれ」という文字列を含む
6	〈雨〉	〈時々雨〉〈一時雨〉以外で「雨」「雪」「みぞれ」という文字列を含む
7	〈時々曇〉	「時々曇」という文字列を含む
8	〈一時曇〉	「一時曇」という文字列を含む
9	〈曇〉	「曇」という文字列と一致する

表 1: 9 分類の定義

と計算する。各分類の定義は表 1 の通りである。

以上の天気予報エントロピーの計算は、予報が全て適中する場合の情報量であるため、実際には予報が外れ、外れた場合に情報量の損失がある。[1] そこで、適中率の持つ情報量も含めて以下のようにエントロピーを導出し検証する。

$$\bar{I} = -p_s \log_2 p_s - p_r \log_2 p_r + p_p \log_2 p_p$$

p_s : その月に降水がなかった確率

p_r : その月に降水があった確率

p_p : 翌日の降水有無の予報適中率

3 解析結果

3.1 47 都道府県

日本 47 都道府県における天気予報エントロピーの月別気候値を 2 分類 (降水の有無)、3,9 分類 (天気カテゴリー) でそれぞれ計算した。地域性を検証するために気象庁が定義している 12 の気候区分を一部まとめて「北海道」「東北」「関東甲信」「北陸」「東海」「近畿」「中国」「四国」「九州」「沖縄」の 10 個の区分にし、各区分で平均を取った。(図 1)

降水の有無の 2 分類で解析を行なった場合、降雪の多い地域で冬にエントロピーが非常に小さくなり、関東甲信・東海・近畿などの内陸部が多い区分では年中通して大きくなるという結果が得られた。(図 1 上段) そのため、降水の多くなる地域・季節で天気予報の必要性が小さくなるという特徴があることが示された。降水の有無は 2 変数であるため、天気予報エントロピーと降水データが同じ変動をしていれば天気予報エントロピーが季節変動を反映していることが確かめられるが、各区分において相関係数を計算した結果、全ての区分で強い相関が見られたため、天気予報エントロピーの有用性を示すことができた。

〈晴れ〉〈雨〉〈曇〉の 3 分類で解析を行なった場合、区分による地域性があまり見られなかったが季節性が大きく反映された。(図 1 中段) どの気候区分も夏・冬

に小さく、春・秋に大きくなっていることから、春・秋の天気予報精度を向上させるべきである。

更に分類を細かくし、〈晴〉〈一時晴〉〈時々晴〉〈雨〉〈一時雨〉〈時々雨〉〈曇〉〈一時曇〉〈時々曇〉の9分類で解析を行なった場合、(1)4,5月に小さく11,12月に大きくなる波の変動(2)夏に大きく冬に小さい山型の変動(3)7月に最小を取り1月に最大を取るV字の変動の3パターンに分けられることが示された。(図1下段)夏に曇や雨になることが多く天気が変わりやすい関東甲信地方が(2)の変動をしていて、夏は晴れの日が多く冬に曇や雨が多く天気が変わりやすい沖縄が(3)の変動をしている。

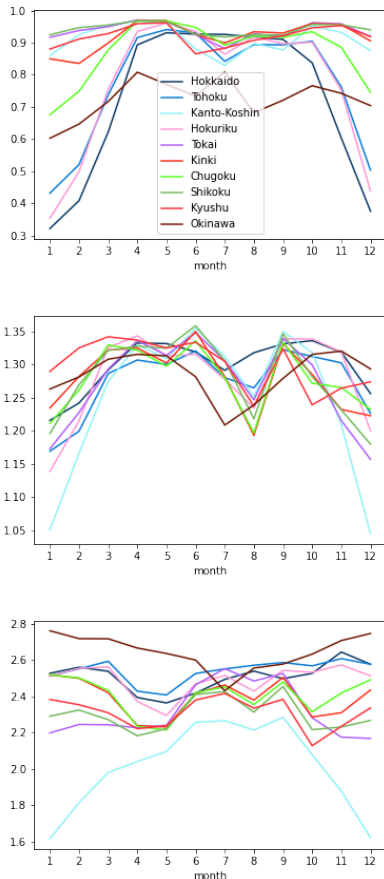


図1: 2分類(上段),3分類(中段),9分類(下段)の天気予報エントロピーの月別気候値

3.2 全球

次に、全球の降水の有無から導出した天気予報エントロピー気候値を描画した(図2)。北半球の冬(DJF)では南極海で天気予報エントロピーが小さくなっている、北半球の夏(JJA)では赤道沿いに天気予報エントロピーの低い線が浮き上がっていて、日本の降水有無の天気予報エントロピーと同様に降水の少ない地域・季節でエントロピーが小さくなっている。

次にエルニーニョ現象やラニーニャ現象との関わりを調べるために、Nino3.4との回帰図を描画した。(図3)エルニーニョ現象が発生すると、西太平洋と東太平洋のエントロピーが高くなり、中央太平洋のエントロピーが低くなる。逆にラニーニャ現象が発生すると、

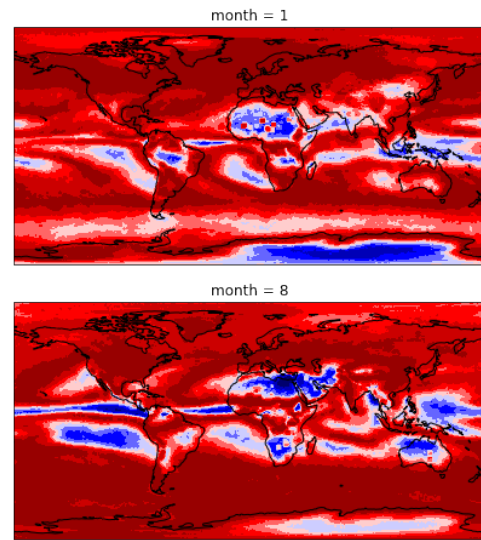


図2: 1月(上段)と8月(下段)の天気予報エントロピー気候値

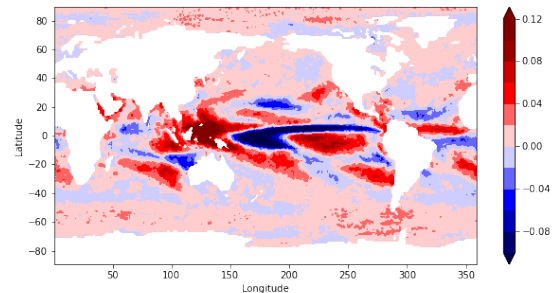


図3: Nino3.4と天気予報エントロピーの回帰図

西太平洋と東太平洋のエントロピーが低くなり、中央太平洋のエントロピーが高くなることが示された。

4 まとめと今後の課題

2分類(降水の有無)においては、ひと月で降水の有無が変化のない季節や地域で天気予報の必要性が低くなるということが定量的に示された。3,9分類(天気カテゴリー)による天気予報エントロピーは分類の数を増やした方が地域性・季節性が大きくなり正確な解析ができることが分かった。全国的には春・秋の天気予報に注力すべきである。また、全球においてはエルニーニョ現象発生時には西太平洋・東太平洋、ラニーニャ現象発生時には中央太平洋に着目して天気予報の精度を上げていくべきだと考えられる。

今後は、3分類・9分類の際も適中率を加味したエントロピーを分析し、全球においても変数を増やして地域性を解析し、より詳細に検証していく必要がある。

参考文献

[1] 正野重方. (1958). 天気予報の適中率. 天気, 5(6)
 [2] Eiichi Suzuki, Weather Forecast and Entropy in Information Theory, 551.509.314, 1958.