階級の予測について

本谷 玲（気象庁気候情報課）

1　序論

この研究では，3 か月予報ガイダンスの改善方法について述べる．

3 か月予報は向こう 3 か月先の平均気温等を，低い，平年並，高いの 3 階級で予報する．3 か月予報ガイダンスは数値予報と階級の関係性を統計的に処理し，数値予報モデルの出力を階級の確率に翻訳するもので，3 か月予報を作成するための参考資料として使用される．よって 3 か月予報ガイダンスの予測精度が向上すれば，3 か月予報の予測精度も改善することが期待できる．

3 か月予報ガイダンスの現行手法は，数値予報から作成された説明変数を，10 メンバーごとに実況の階級と線形回帰している．この現行手法には問題があることを説明し，統計的機械学習の観点から現行手法の問題点を修正する．最後に現行手法と修正した手法の予測精度の対照実験を行い，修正した手法の予測精度が大きく優越することを示した．すなわちこの研究は，長期予報の予測精度を統計的機械学習でもって改善する事例となっている．

2　説明変数の作成

数値予報から説明変数 $\mathbf{x}$ を生成する．対象時刻 t，リードタイム l，メンバー m，300hPa 高度などの数値予報の予測対象を表す添字 d，等緯度経度格子の番号 i, j の予報値を x_{tlmdij} とする．$x^{\mathrm{climate}}_{tdij}$ はモデル気候値．この資料に記載した実験では，現行ルーチンのモデル気候値を使用した．α_{ij} は等緯度経度の格子ごとの緯度の違いによる面積比を表現した重み係数．R^{region} は地域 region の地域平均に用いる格子点の集合で，CPS2 の場合，遠藤らの図 1.6.1 に対応する [4]．$T^{\mathrm{year,month}}$ は year, month を 3 か月予報の初月とする，3 か月予報の予報期間となる対象時刻の集合．T'^{month} はモデル気候値における $T^{\mathrm{year,month}}$ と同等の期間．

$$x^{\mathrm{region,year,month}}_{lmd} = \frac{1}{|T^{\mathrm{year,month}}|}\sum_{t\in T^{\mathrm{year,month}}}\sum_{(i,j)\in R^{\mathrm{region}}}\alpha_{ij}x_{tlmdij} - \frac{1}{|T'^{\mathrm{month}}|}\sum_{t\in T'^{\mathrm{month}}}\sum_{(i,j)\in R^{\mathrm{region}}}\alpha_{ij}x^{\mathrm{climate}}_{tdij}. \quad (1)$$

$$\mathbf{x}^{\mathrm{region,year,month}}_{lm} = \left(x^{\mathrm{region,year,month}}_{lm1}, \ldots, x^{\mathrm{region,year,month}}_{lmD}\right)^{\top}. \quad (2)$$

CPS2 ハインドキャストから 3 か月の予報期間を切り取る場合，5 メンバー ×2LAF でリードタイム 0 か月〜4 か月の 5 通りのサンプリングができる．同様に CPS3 ハインドキャストから 3 か月の予報期間を切り取る場合，5 メンバー ×2LAF でリードタイム 0 か月〜6 か月の 7 通りのサンプリングができる．

3　現行手法

現行手法の方法に従い，説明変数から確率予測を行う [4]．数値予報により生成された説明変数 $\mathbf{x}_{nm} \in \mathbb{R}^D$ と，実況の平年差もしくは平年比 y_n の組の集合 $\mathcal{D} = \{\langle \mathbf{x}_{nm}, y_n\rangle\}$ が与えられている．N はサンプル数，M はメンバー数，D は説明変数の次元数．

適当なメンバー数 $M_2 = 10$ ごとに平均をとる．これを肩に $^+$ を付けて表現することにする．

$$\mathbf{x}^{+}_{nm_1} = \frac{1}{M_2}\sum_{m_2}\mathbf{x}_{n,M_2(m_1-1)+m_2}. \quad (3)$$

尤度関数 L からパラメタ $\hat{\mathbf{w}}, \hat{b}, \hat{\sigma}^2$ を定める．$\mathcal{N}(\cdot,\cdot)$ は正規分布．

$$L(\mathbf{w}, b, \sigma^2, \{\langle \mathbf{x}^{+}_{nm_1}, y_n\rangle\}) = \prod_{n,m_1}\mathcal{N}(y_n \mid \mathbf{w}^{\top}\mathbf{x}^{+}_{nm_1} + b, \sigma^2). \quad (4)$$

$$\hat{\mathbf{w}}, \hat{b}, \hat{\sigma}^2 = \arg\max_{\mathbf{w},b,\sigma^2} L(\mathbf{w}, b, \sigma^2, \{\langle \mathbf{x}^{+}_{nm_1}, y_n\rangle\}). \quad (5)$$

遠藤らは情報量基準 AICc に従い，$\mathcal{F}$ から学習データを使って説明変数を選択しているので，現行手法では説明変数の選択にこれを使う．[*1]

$$\mathrm{AICc} = N(\ln 2\pi\hat{\sigma}^2 + 1) + \frac{2(D+2)N}{N-(D+2)-1}. \quad (6)$$

M_{valid} 個の未知の説明変数 $\{\mathbf{x}'_m\}$ についての確率予測を得る．t_{k1}, t_{k2} は階級 k の階級区分値．

$$p(\mathcal{C}_k \mid \hat{\mathbf{w}}, \hat{b}, \hat{\sigma}^2, \{\mathbf{x}'^{+}_{m_1}\}) = \frac{1}{\lfloor M_{\mathrm{valid}}/M_2\rfloor}\sum_{m_1}\int_{t_{k2}}^{t_{k1}}\mathcal{N}(t \mid \hat{\mathbf{w}}^{\top}\mathbf{x}'^{+}_{m_1} + \hat{b}, \hat{\sigma}^2)\,\mathrm{d}t. \quad (7)$$

3.1　現行手法の問題点 1 ：不自然な前提

現行手法は数値予報の各メンバーの順序に意味があることを前提としている．$N = 1$ に限定し，$\{\mathbf{x}_m\}$ について考える．

[*1] 真の分布と予測分布が同じ族であるという前提が成り立つか不明であるため，本来この問題に AICc を使うことはできない [1]．しかし遠藤らは AICc を使っているから，ひとまず現行手法での説明変数の選択では AICc を利用した．

$\boldsymbol{\theta} = \langle \mathbf{w}, b, \sigma^2 \rangle$. $a \in \{1, \ldots, M_2\}, b \in \{M_2+1, \ldots, 2M_2\}$.

$$\mathbf{x}_1^+ = \frac{1}{M_2}\sum_{m_2} \mathbf{x}_{m_2}, \qquad \mathbf{x}_2^+ = \frac{1}{M_2}\sum_{m_2} \mathbf{x}_{M_2+m_2},$$
$$\mathbf{x}_3^+ = \mathbf{x}_1^+ - \mathbf{x}_a/M_2 + \mathbf{x}_b/M_2, \quad \mathbf{x}_4^+ = \mathbf{x}_2^+ + \mathbf{x}_a/M_2 - \mathbf{x}_b/M_2. \tag{8}$$

これらについて一般に

$$L(\boldsymbol{\theta}, \langle \mathbf{x}_1^+, y\rangle, \langle \mathbf{x}_2^+, y\rangle) \neq L(\boldsymbol{\theta}, \langle \mathbf{x}_3^+, y\rangle, \langle \mathbf{x}_4^+, y\rangle) \tag{9}$$

だから，de Finetti の表現定理より現行手法は $\{\mathbf{x}_m\}$ を独立同分布ではないことを前提としている．

つまり現行手法は，数値予報のメンバーの順序と現実の階級の間に，なんらかの関係を想定している．例えば CPS2 ルーチンは 13 メンバー ×4LAF だが，13 メンバーの内，1 番目のメンバーと 13 番目のメンバーの順序を入れ替えると現行手法の出力する確率予測は変化してしまう．しかしながら，1 番目のメンバーと 10 番目のメンバーの順序を入れ替えても，現行手法の出力する確率予測は変化しない．

よって 10 番目と 13 番目のメンバーの取り扱いの違いに有効な説明が存在する場合に限定して，すなわち数値予報の各メンバーは独立同分布ではなく，なおかつ 10 メンバー単位では交換可能という，不自然な前提が成立する場合に限り，現行手法は統計的な合理性を持つ．しかし遠藤らは数値予報の各メンバーの順序による取り扱いの差異について説明していないし，それを気にしているような記述もない [4]. 10 メンバーずつ平均をとって線形回帰するという現行手法は，統計的な合理性を意識せずに設計されたように思われる．

また 10 メンバーずつ平均をとる操作によって，本来数値予報のメンバーがもつ確率分布の情報を喪失させている点も問題である．

3.2 現行手法の問題点２：非効率な手法

現行手法は初めに平年差や平年比を予測して，平年差の予測からさらに階級の確率を予測するという二段階の予測になっている．しかしながら，3 か月予報について平年差の予測ではなく階級の予測を発表しているのは，階級の予測の方が比較的簡単だと考えられたからだ．現行手法は平年差を予測するという相対的に困難な問題を解くことで，階級の確率を予測するという緩和された問題を解決しようとしている．

4 修正した手法

先述した現行手法の問題点を，統計的機械学習の枠組みで修正する．数値予報により生成された説明変数 $\mathbb{R}^D \ni \mathbf{x}_{nm} \overset{\text{i.i.d.}}{\sim} F_n(\mathbf{x}_n)$ と，現実での階級を表現する 1-of-K ベクトル $\mathbf{c}_n \in \{0,1\}^K$ の組の集合 $\mathcal{D} = \{\langle \mathbf{x}_{nm}, \mathbf{c}_n\rangle\}$ が与えられている．N はサンプル数，M はメンバー数，K は階級の数，D は説明変数の次元数．以下が成立する [2].

$$F_{nM}(\mathbf{x}_n) = \frac{1}{M}\sum_m \mathbf{1}(\mathbf{x}_{nm} \leq \mathbf{x}_n) \xrightarrow{\text{a.s.}} F_n(\mathbf{x}_n). \tag{10}$$

$\mathbf{x}_n^* \overset{\text{i.i.d.}}{\sim} F_{nM}(\mathbf{x}_n)$ とする．$\mathcal{U}(\cdot)$ は一様分布．

$$\boldsymbol{\mu} = \mathbb{E}\left[\frac{1}{N}\sum_n \mathbf{x}_n^*\right]. \tag{11}$$

$$\boldsymbol{\sigma}^2 = \mathbb{E}\left[\frac{1}{N}\sum_n (\mathbf{x}_n^* - \boldsymbol{\mu}) \odot (\mathbf{x}_n^* - \boldsymbol{\mu})\right]. \tag{12}$$

$$p(\mathbf{x}_n) \simeq \mathcal{U}\left(\{(\mathbf{x}_n^* - \boldsymbol{\mu}) \oslash \boldsymbol{\sigma}\}\right). \tag{13}$$

階級 $\mathcal{C}_k$ の確率予測 $p(\mathcal{C}_k \mid \boldsymbol{\theta}, \mathbf{x}_n)$ の KL ダイバージェンス E から，$\hat{\boldsymbol{\theta}}$ を学習できる．

$$F(t \mid \alpha) = \frac{1}{1+\exp(\alpha - t)}. \tag{14}$$

$$p(\mathcal{C}_k \mid \boldsymbol{\theta}, \mathbf{x}_n) = F(t_{k1} \mid \boldsymbol{\theta}^\top \mathbf{x}_n) - F(t_{k2} \mid \boldsymbol{\theta}^\top \mathbf{x}_n). \tag{15}$$

$$\begin{aligned} E(\boldsymbol{\theta}, \{\langle \mathbf{x}_n, \mathbf{c}_n\rangle\}) &= D_{\mathrm{KL}}(\mathbf{c}_n \,\|\, p(\mathcal{C}_k \mid \boldsymbol{\theta}, \mathbf{x}_n)) \\ &\propto -\frac{1}{N}\sum_{n,k} c_{nk} \ln p(\mathcal{C}_k \mid \boldsymbol{\theta}, \mathbf{x}_n) + \text{const.} \end{aligned} \tag{16}$$

$$\hat{\boldsymbol{\theta}}(\{\langle \mathbf{x}_n, \mathbf{c}_n\rangle\}) = \arg\min_{\boldsymbol{\theta}} E(\boldsymbol{\theta}, \{\langle \mathbf{x}_n, \mathbf{c}_n\rangle\}). \tag{17}$$

$\hat{\boldsymbol{\theta}}$ から，未知の説明変数 $\{\mathbf{x}'_m\}$ についての確率予測が得られる．

$$\begin{aligned} \mathbb{E}[p(\mathcal{C}_k \mid \boldsymbol{\theta}, \mathbf{x}')] = \int & p(\mathcal{C}_k \mid \hat{\boldsymbol{\theta}}(\{\langle \mathbf{x}_n, \mathbf{c}_n\rangle\}), \mathbf{x}') p(\mathbf{x}') \\ & \prod_n p(\mathbf{x}_n)\, \mathrm{d}\mathbf{x}' \mathrm{d}\mathbf{x}_1 \ldots \mathrm{d}\mathbf{x}_N. \end{aligned} \tag{18}$$

修正した手法は数値予報のメンバーが独立同分布に従うことを前提として導出されているため，メンバーの順序を入れ替えても予測結果は変わらない．修正した手法は直接的に各階級の確率を予測するので，平年差を予測するというより困難な問題を解くことを回避している．

5 実験

現行手法と修正した手法の予測精度を比較し，修正した手法の優越性を示す．3 か月平均気温について，CPS2 及び CPS3 ハインドキャストを使って現行手法と修正した手法で 3 階級を確率予測させたときの Brier skill score (BSS) を比較した [3]. Nested cross validation (Nested CV) により，学習データと検証データを分離し，学習データだけを使って，3 か月平均気温の説明変数の候補 $\mathcal{F}$ からどの説明変数を使うかのハイパーパラメタと，学習パラメタを決定した [5]. 表 1 に実験の設定をまとめる．検証の手順を Algorithm 1, Algorithm 2 にそれぞれ示す．

$$\mathcal{F} = \begin{cases} \{\mathrm{z500}, \mathrm{z300}, \mathrm{ts}, \mathrm{t850}\} & \text{CPS2 の場合,} \\ \{\mathrm{z850}, \mathrm{z500}, \mathrm{z300}, \mathrm{ts}, \mathrm{t925}, \mathrm{t850}\} & \text{CPS3 の場合.} \end{cases} \tag{19}$$

実験結果を図 1 に示す．現行手法のままでも，CPS2 から CPS3 への改良により BSS を改善している．一方で CPS3 での修正した手法の BSS は，明らかに CPS3 での現行手法の BSS を優越している．これは数値予報は十分な情報量を持っているにもかかわらず，現行手法がそれをスポイルしてしまっていることを意味している．修正した手法はこれを改善し，予測精度を向上させた．

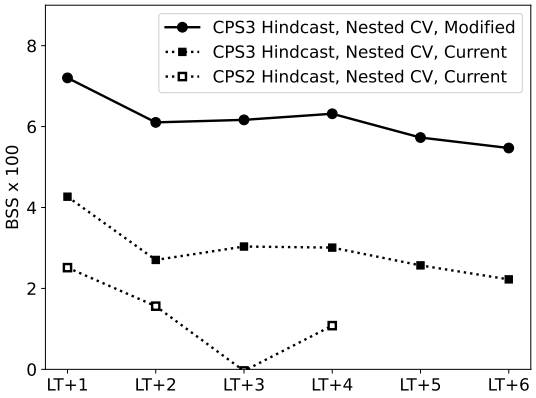

図 1　3 か月平均気温を 3 階級で予測したときの予測精度．横軸はリードタイムで，例えば LT+1 は予測をするのに使った数値予報のリードタイムが 1 か月であることを意味する．縦軸は 3 階級で予測したときの各地域の BSS の平均値を 100 倍した値で，0 のとき気候学的確率による予測と同じ予測精度，100 のとき完全予報．折れ線グラフは上から順に，修正した手法 (CPS3)，現行手法 (CPS3)，現行手法 (CPS2) をそれぞれ表す．

6　結論

この研究では，

(1) 3 か月予報ガイダンスの現行手法は統計学を無視したものになっており，数値予報をスポイルしていること，
(2) これまでスポイルされていた情報は統計的機械学習の枠組みで活用できること，
(3) これにより予測精度を圧倒的に改善すること，

の 3 点を実証した．

3 か月予報以外の季節予報のガイダンスや他の発表情報についても，同じように数値予報をスポイルしているから，この枠組みで改善が可能であろう．

参考文献

[1] Clifford M. Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.
[2] Michael Naaman. On the tight constant in the multivariate Dvoretzky-Kiefer-Wolfowitz inequality. *Statistics & Probability Letters*, 173, 2021.
[3] 小林 大輝. 付録 A 数値予報課報告・別冊で用いた表記と統計的検証に用いる代表的な指標. **数値予報課報告・別冊**, 64, 2018.
[4] 遠藤 新 and 竹川 元章. 1.6 3 か月予報および暖・寒候期予報のためのガイダンスの更新. **季節予報研修テキスト**, 28, 2015.
[5] 工藤 淳. 2.3 ガイダンスに用いる統計の基礎. **数値予報課報告・別冊**, 64, 2018.

表 1　3 か月平均気温のガイダンスの予測精度を評価するための実験の設定．

	現行手法（CPS2）	現行手法（CPS3）	修正した手法（CPS3）
使用したデータ	CPS2 ハインドキャストから生成した説明変数のうち，予報初月 1991 年 1 月から 2020 年 12 月，リードタイム 1 か月〜4 か月のデータを使った．	CPS3 ハインドキャストから生成した説明変数のうち，予報初月 1991 年 1 月から 2020 年 12 月，リードタイム 1 か月〜6 か月のデータを使った．予報対象月の前後の月を学習データに加えている．	
平年差及び階級区分値	2020 年平年値の平年差及び階級区分値を使った．		
検証方法	Nested CV により学習データと検証データを分離して，学習データのみからパラメタやハイパーパラメタを推定し，各手法の予測精度を計算した．		
検証する地域	北海道，東北，関東甲信，北陸，東海，近畿，中国，四国，九州北部，九州南部，奄美，沖縄の各地方．（reg. 11, 15, 20, 21, 22, 23, 26, 29, 30, 32, 33, 34）		
検証指標	検証する地域ごとに 3 か月平均気温を 3 階級で予測させ，そのときの BSS の平均値を評価する．		

Algorithm 1 Nested cross validation of the current method

for region **do**
  for year **do**
    for month $\in \{1, \ldots, 12\}$ **do**
      for factor $\in \mathcal{F}$ **do**
        $\mathcal{D}_{\text{train}}^{\text{factor,year}} \leftarrow \mathcal{D}_{n\neq\text{year},m}^{\text{factor,region,month}}$.
        compute $\text{AICc}(\mathcal{D}_{\text{train}}^{\text{factor,year}})$.
      end for
      $\widehat{\text{factor}} \leftarrow \arg\min_{\text{factor}} \text{AICc}(\mathcal{D}_{\text{train}}^{\text{factor,year}})$.
      $\mathcal{D}_{\text{train}}^{\widehat{\text{factor}},\text{year}} \leftarrow \mathcal{D}_{n\neq\text{year},m}^{\widehat{\text{factor}},\text{region,month}}$.
      $\mathcal{D}_{\text{valid}}^{\widehat{\text{factor}},\text{year}} \leftarrow \mathcal{D}_{n=\text{year},l=1,m}^{\widehat{\text{factor}},\text{region,month}}$.
      $\hat{\mathbf{w}}, \hat{b}, \hat{\sigma}^2 \leftarrow \arg\max_{\mathbf{w},b,\sigma^2} L(\mathbf{w}, b, \sigma^2, \mathcal{D}_{\text{train}}^{\widehat{\text{factor}},\text{year}})$.
      validate $p(\mathcal{C}_k \mid \hat{\mathbf{w}}, \hat{b}, \hat{\sigma}^2, \mathcal{D}_{\text{valid}}^{\widehat{\text{factor}},\text{year}})$.
    end for
  end for
end for

Algorithm 2 Nested cross validation of the modified method

for region **do**
  for year **do**
    for month $\in \{1, \ldots, 12\}$ **do**
      for factor $\in \mathcal{F}$ **do**
        for year$'$ $\neq$ year **do**
          for leadtime **do**
            $\mathcal{D}_{\text{train}}^{\text{factor,year}',\text{leadtime}} \leftarrow \mathcal{D}_{n\neq\text{year}\wedge n\neq\text{year}',l=\text{leadtime},m}^{\text{factor,region,month}}$.
            $\mathcal{D}_{\text{valid}}^{\text{factor,year}',\text{leadtime}} \leftarrow \mathcal{D}_{n\neq\text{year}\wedge n=\text{year}',l=\text{leadtime},m}^{\text{factor,region,month}}$.
            compute $\mathbb{E}[E(\hat{\boldsymbol{\theta}}(\mathcal{D}_{\text{train}}^{\text{factor,year}',\text{leadtime}}), \mathcal{D}_{\text{valid}}^{\text{factor,year}',\text{leadtime}})]$.
          end for
        end for
      end for
      $\widehat{\text{factor}} \leftarrow \arg\min_{\text{factor}} \mathbb{E}_{\text{year}',\text{leadtime}} \mathbb{E}\left[E(\hat{\boldsymbol{\theta}}(\mathcal{D}_{\text{train}}^{\text{factor,year}',\text{leadtime}}), \mathcal{D}_{\text{valid}}^{\text{factor,year}',\text{leadtime}})\right]$.
      $\mathcal{D}_{\text{train}}^{\widehat{\text{factor}},\text{year}} \leftarrow \mathcal{D}_{n\neq\text{year},l=1,m}^{\widehat{\text{factor}},\text{region,month}}$.
      $\mathcal{D}_{\text{valid}}^{\widehat{\text{factor}},\text{year}} \leftarrow \mathcal{D}_{n=\text{year},l=1,m}^{\widehat{\text{factor}},\text{region,month}}$.
      validate $\mathbb{E}\left[p(\mathcal{C}_k \mid \hat{\boldsymbol{\theta}}(\mathcal{D}_{\text{train}}^{\widehat{\text{factor}},\text{year}}), \mathcal{D}_{\text{valid}}^{\widehat{\text{factor}},\text{year}})\right]$.
    end for
  end for
end for