生成AIを活用した需要予測は企業の意思決定にどこまで貢献できるか

― モデルベース型時系列予測手法との比較にみる非連続的需要構造への適用可能性と在庫戦略への示唆 ―

山崎　倫未

キーワード：生成AI、需要予測、Prophet、非連続的需要構造、釣具需要、意思決定支援

1．はじめに

近年、多くの製造業で需要の不確実性が高まり、在庫管理や生産計画の難易度が増している。釣具製造業もその例外ではなく、天候や海況、社会的イベントなど多様な外生要因に加え、円安による原材料高騰や低価格ECチャネル拡大による価格競争も進んでおり、需要予測の高度化は重要な経営課題となっている。

一方、需要予測に関する既存の理論的知見の多くは、電力需要予測に代表されるように、需要の変動を連続的な量の増減として捉えることを前提として蓄積されてきた。しかし、釣具製造業における定番製品「サビキ」のように、「釣りに行くか否か」という離散的な行動選択そのものが需要発生の前提となる財については、こうした既存の理論的枠組みがどこまで適用可能かは、2-1 節で詳述する通り十分に検証されていない。

この理論的空白は、対象企業である釣具製造業 A 社が抱える実務上の課題とも結びついている。同社でも、定番製品「サビキ」を中心に、需要予測の多くが担当者の経験や勘に依存し、多大な作業時間を要するとともに外生要因を反映しにくいという課題

を抱えており、これは、離散的な需要構造を扱うための理論的な蓄積が乏しいことの一つの表れと捉えることができる。

本研究は、こうした理論上・実務上の課題を踏まえ、近年、生成AIの企業活動における活用が徐々に進みつつある（山口ほか，2024）流れの中で、需要予測という業務プロセスへ生成AI適用事例を提供するものである。

本研究の目的は、生成AIを用いた需要予測が、従来のモデルベース型時系列予測手法であるProphetと比較して企業の意思決定にどのように貢献し得るのかを、生成AIを自動化手段ではなく経験や勘を支援・補完する意思決定支援ツールと位置づけた上で明らかにすることである。

本稿の構成は次の通りである。２節では先行研究を踏まえたリサーチギャップを、３節では対象企業の現状と課題構造を、４節では本研究の目的・リサーチクエスチョン・分析設計を、５節ではProphet単体と生成AI補完後の予測の構築・比較を、６節では結論、実務的・学術的意義、限界と今後の課題を、それぞれ述べる。

２．需要予測に関する先行研究とリサーチギャップ

2-1．先行研究の整理

需要予測研究は、電力・小売・製造など多様な分野で蓄積されてきたが、対象とする需要の性質によって重視される課題は異なる。安田ほか（2021）は商品特性によって予測難易度が異なることを、永井ほか（2025）はデータが少ない商品ほど商品属性等の外部情報の活用が有効であることを、森山・井對（2025）は時系列分解と外部データの組み合わせが説明性の高い予測を可能にすることを、それぞれ示している。一方、電力需要予測では、渡邉ほか（2020）、中山・浦野（2023）、Xie（2016）が示すように、高頻度データや気象情報を組み込んだモデルベース型手法が高度に蓄積されている。ただし、Prophetに代表される時系列分解の枠組みも、こうした電力需要予測のモデルベース型手法も、需要の変動を連続的な量の増減として捉えることを前提としている。これに対し、サビキ需要のように「釣りに行くか否か」という離散的な行動選択そのものが需要発生の前提となる製品では、変動が量的な多寡ではなく、発生するかしないかという質的な分岐に由来する可能性がある。したがって、これらの知見が、サビキ需要のような非連続的かつレジャー依存的な需要構造にどこまで理論的に適用可能かは、十分に検証されていない。なお、担当者の経験や勘への依存や作業負担といった実務上の課題は１節で述べた通りであり、ここで指摘する適用可能性の問題は、それとは

別の理論的な課題である。

なお、受注ゼロの月が多く含まれるような、ゼロ過多で断続的に発生する需要系列を扱う専門領域として、間欠需要予測（intermittent demand forecasting）が知られ、Croston 法（Croston, 1972）やその改良である TSB 法（Teunter et al., 2011）などが提案されている。ただし本研究は、これら間欠需要専用手法そのものの改良を目的とするのではなく、Prophet という連続的な量の増減を前提とするモデルに生成 AI の定性的判断を組み合わせる設計の有効性を検証することに主眼を置く。後述するゼロ予測の壁という現象は、こうした間欠需要予測の領域と接続する課題として位置づけられるが、専用手法との直接比較は本研究の範囲を超えるため、6-5 節で述べる今後の課題とする。

2-2. モデルベース型時系列予測手法（Prophet）

Prophet（Taylor and Letham, 2018）は、トレンド・季節性・休日効果を加法モデルとして分解し、解釈可能な形で時系列構造を捉える手法であり、実務での運用のしやすさから広く用いられている。本研究では、サビキ需要の時系列構造を可視化する基準モデルとして Prophet を採用する。ただし、Prophet は天候・イベント・商品属性といった定性的要因を柔軟に取り込む仕組みを有しておらず、コロナ特需のような一時的な特殊局面を時系列構造の一部として単純に吸収できるとは限らない。また、弓岡（2023）が指摘するように学習データが少ない場合には精度が低下しやすいという限界もある。

なお、XGBoost や LSTM 等、より高精度な手法も存在するが、本研究では、判断根拠を担当者が直感的に理解できる解釈可能性と、専門人材を要しない実務導入の容易さを重視し、Prophet を比較対象として採用した。これは単なる方針上の選好ではなく、本研究の比較設計上の要請でもある。本研究の目的は、生成 AI が基準モデルの数値に対して「どのような理由で」補正を加えたのかを追跡可能な形で検証することにあり、そのためには、基準モデルの予測値自体がなぜその値になったのかを説明できることが前提となる。Prophet はこうした明示的な構造を持つため、AI が提案する補正を、この既知の構造に対する追加的な調整として位置づけることができる。これに対し、XGBoost や LSTM のような手法は、予測値が多数の特徴量の非線形な相互作用から導かれるため、基準モデルの予測値自体がなぜその水準になったのかの説明が難しく、AI による補正が生成 AI 自体の定性的判断に起因するのか、基準モデルの複雑な内部構造に起因するのかを切り分けられなくなる。Prophet の解釈可能性は、この

追跡可能性を確保するための前提条件である。他手法との比較は、5-5 節の今後の課題とする。

2-3. 生成 AI を用いた需要予測手法

生成 AI は、自然言語による定性情報を統合的に解釈し、根拠を伴う説明を生成できる点で、Prophet の数値結果だけでは伝わりにくい「なぜこの予測になるのか」という問いに答える役割を果たし得る。本研究では、Prophet の予測結果に天候・休日・商品属性等の定性情報を与え、補正の方向性を提案させることで、後述する 3-1 節で述べる「担当者の経験や勘による補正」を支援することを想定する。ただし、生成 AI の出力は必ずしも事実に基づいた正確なものとは限らず、いわゆるハルシネーションを起こす可能性があるという限界がある。また、伊藤・中川（2024）が指摘するように、プロンプトの与え方によって出力が変動しやすいという限界もある。したがって、生成 AI は Prophet の代替手法ではなく、その数値結果を前提とした補完的な解釈・提案の手段として位置づける。

2-4. リサーチギャップ

2-1 節から 2-3 節で確認した通り、電力需要のような安定的・連続的需要に関する知見、Prophet の解釈可能性とその限界、生成 AI の定性的解釈能力とその限界は、それぞれ個別には蓄積されている。しかし、これらを組み合わせても、次の３つのギャップは理論的に埋まらない。第一に、連続的需要を前提とする知見が、需要発生自体が非連続的な構造にどこまで転用可能かは理論的枠組みからは導き出せず、電力需要予測で蓄積された知見が、非連続的かつレジャー依存的な需要構造へ適用が可能かどうかについては、十分に検証されていない。第二に、Prophet と生成 AI はそれぞれ独立に評価されてきたに過ぎず、両者を補完的に組み合わせた場合の効果の大きさやその条件は個々の手法の特性から自動的には導かれないため、両者を補完的に組み合わせその効果を定量検証した研究は乏しい。第三に、たとえ精度が改善したとしても、それが企業の意思決定にどのような価値をもたらすのかは精度指標だけでは説明できず、予測結果が企業の意思決定にどのように貢献するのかは十分に検討されていない。

これらのギャップを検証するには、非連続的かつレジャー依存的な需要構造を実際に持つ、具体的な事例による検証が必要となる。次節では、この観点から対象企業を選定した経緯を述べる。

３．現状分析と課題の構造

3-1. 対象企業における現状と課題

２節で示したギャップを検証するため、本研究では釣具製造業 A 社のサビキ需要を対象とする。同社の需要予測の現状を把握するため、筆者は 2026 年４月 22 日、同社釣具部門責任者への半構造化インタビューを実施した。インタビューによれば、同社の需要予測は、過去約７年分の販売実績データとトレンドグラフを基礎とし、担当者の経験や勘による補正を加えて翌年度１年分の予測を１件１件確認して確定している。この作業はExcel 上での手作業を前提とするため多大な時間と労力を要する。加えて、同責任者は、天候や社会的イベントといった外的要因を予測により反映させたいという意向を持ちながらも、現状の手作業を前提とする体制ではその余裕がないこと、また、本来であれば年度内に複数回、予測を見直したいという希望があるものの、１件１件の確認作業に要する時間の制約から、年１回の見直しにとどまらざるを得ないことを述べた。さらに同責任者は、BtoB 構造による最終需要の間接的な把握、季節性・天候への強い依存、需要の急激な変化といった要因から、自社製品の需要予測は本質的に難しいのではないかという認識を示した。なお、本研究では、計上ルールや在庫制約の影響を受けにくい受注データを、需要をより直接的に反映するデータとして採用する。

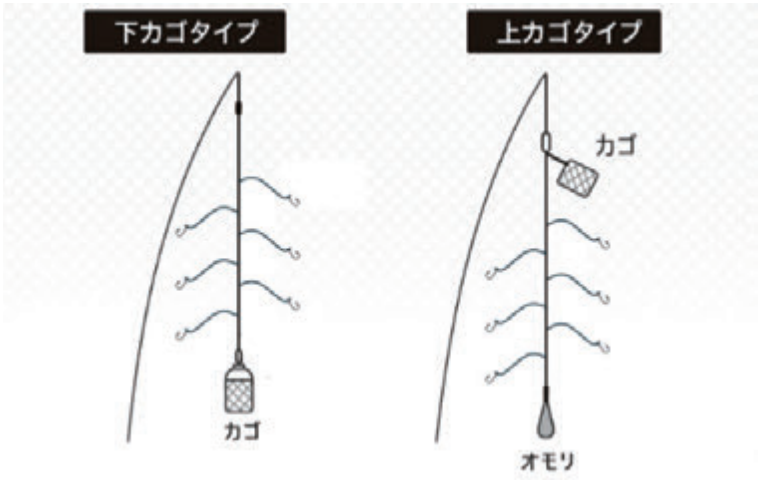

図１：サビキ仕掛けの基本構造

本研究では、同社の定番製品であるサビキを分析対象とする。サビキの基本構造を

図１に示す。サビキは仕様の違いにより複数の SKU[1]に細分化されており、本研究では SKU を予測・在庫管理の最小単位として扱う。

3-2. 業界構造・意思決定構造とサビキの需要特性

釣具業界は、消耗品中心の製品特性ゆえに新規参入・代替品の脅威が高く、低価格チャネルの拡大により価格競争が常態化している。同社を含む多くの釣具メーカーは BtoB 販売を基本とし、最終消費者の購買行動を直接把握できないため、買い手である流通事業者の交渉力が相対的に強い。加えて、海外生産に伴う調達リードタイムの長さから、需要予測の誤差が在庫リスクとして顕在化しやすい。また、釣り場・堤防への立入規制強化による釣行機会の減少、円安による原材料費上昇と価格感応度の高まり、釣り人口の高年齢化や余暇活動の多様化による需要分散の拡大、EC プラットフォームの普及に伴う消費者購買データの間接化など、政治・経済・社会・技術面の外部環境要因が需要予測を複雑にしている。これらは相互に作用しながら需要の水準や変動幅に影響を及ぼしており、釣具需要では、これらの要因が需要の発生自体に影響する点が特徴的である。

同社の意思決定構造を確認すると、同社は BtoB 販売を基本とし最終消費者への直接販売を行わないため需要情報を取引先経由でしか把握できず、海外工場での生産から在庫化までのリードタイムの長さから需要変動に応じた短期的な生産調整も難しい。消耗品分野での価格競争の常態化により、需要予測は「当てること」に加えて「外れた場合のリスク」を考慮した在庫判断材料としても利用されている。同社の意思決定プロセスは、SKU 単位での実績・トレンド確認、経験や勘による補正、概ね毎年５月頃における翌年度予測の確定、生産・在庫計画への反映、という流れで構成され、需要予測は単独で完結せず、生産・在庫・資金に関する意思決定と密接に結びついている。

サビキは、「釣りを行うか否か」という行動特性そのものが需要発生の前提となる消耗品であり、堤防釣り・船釣りといった用途別に市場が細分化されている。価格の需要への影響は限定的であり、価格よりも釣行機会の有無や製品仕様の適合性が需要を左右する。同社は低価格訴求ではなく、釣果の再現性・信頼性を軸としたポジショニングを確立しており、サビキは定番製品として中長期的な需要基盤を持つ一方、SKU 単位では仕様差により需要が大きく分散する。加えて、天候や海況の影響には地域差があり、波が高く船が出航できない場合には釣行自体が中止となり需要が大きく減少するが、

[1] SKU（Stock Keeping Unit）：在庫管理・受発注管理上の最小識別単位。A 社では、１つの品番に対し、針のサイズや糸（ハリス）の号数などの違いにより複数の SKU が存在する。

その程度は地域の気象・海況条件によって異なる。また、コロナ禍における屋外レジャー需要の急増とその後の反動減のように、通常の季節性やトレンドでは説明できない一時的な特殊局面も存在した可能性がある。こうした地域差や特殊要因は、Prophet ではなく生成 AI に定性的な文脈情報として与えることが妥当と考えられる。

以上のように、釣具業界の競争構造、外部環境、対象企業の意思決定構造、サビキ製品の需要特性はいずれも、需要が安定的・連続的に発生する電力業界とは対照的な性質を持つ。この対照関係を表１に示す。なお、表１における「釣具業界（サビキ需要）」の欄は、PEST および５Forces については業界全体に共通する構造的特徴を示す一方、3C（意思決定構造）については、BtoB 構造という業界共通の制約が同社において具体的にどのように意思決定プロセスへ顕在化しているかを示す事例であり、他の釣具メーカーの意思決定プロセスを直接一般化するものではない。また、STP／4P における用途・仕様ごとの需要分散は、サビキという製品カテゴリーに固有の特性であり、釣具業界全体の需要特性を必ずしも代表するものではない。

表１：釣具業界（サビキ需要）と電力業界の需要構造比較

観点	釣具業界（サビキ需要）	電力業界（比較対象）
PEST（外部環境）	規制・天候・社会イベントが「需要の発生自体」に影響	政治・経済・社会・技術要因は主に「需要の水準・変動幅」に影響
５Forces（競争構造）	価格競争が常態化、最終需要を間接的にしか把握できない	生活インフラとして需要は安定的・連続的
３C（意思決定構造）	需要データを取引先経由でしか取得できない BtoB 構造（A 社の場合、生産・在庫・資金の意思決定に直結）	需要データを高頻度・直接的に取得可能、短期的な需給管理が中心
STP/４P（需要特性）	レジャー行動依存、用途・仕様ごとに需要が分散	全消費者対象の単一・均質な需要

サビキ需要は、定番製品としての継続的需要基盤と SKU 単位での大きな変動が併存し、最終需要を流通事業者経由でしか把握できず予測誤差が生産・在庫・資金の意思決定に直結し、外生要因や地域差・特殊局面によって複合的に変動し、商品属性による需要差が大きく、現行予測が属人的かつ多大な時間を要するという特徴を持つ。これら

は、モデルベース型時系列予測手法と生成 AI という異なる特性を持つ手法を比較検討する対象として、サビキ需要が学術的にも妥当性を持つことを示している。ただし、両者が対照的な性質を持つことは、電力需要予測の知見がサビキ需要に一切適用できないことを意味しない。表１に示した対照関係は、需要がいつ・どの程度発生するかを規定する構造上の違いを整理したものであり、いずれの知見が有効か無効かをあらかじめ確定するものではない。むしろ、こうした構造的な違いが明確であるからこそ、電力需要予測の知見をそのまま適用した場合にどこで機能し、どこで機能しなくなるのかを、実際のデータを用いて検証する意義がある。なお、こうした構造的な予測困難性は、3-1 節で述べた同責任者自身の認識、すなわち BtoB 構造・季節性や天候への依存・需要の急激な変化ゆえに自社製品の需要予測が本質的に難しいという実感とも一致している。この一致は、筆者による分析が、当事者へのインタビューを通じて得られた実感によっても裏付けられていることを示している。

こうした対照性と、そこから見出される検証の意義を踏まえると、２節で示した３つのギャップ、すなわち非連続的・レジャー依存的な需要構造への電力需要予測知見の適用可能性、Prophet と生成 AI の補完的活用の実証、需要予測の意思決定支援としての価値のいずれについても、釣具製造業 A 社のサビキ需要は、これらを検証する事例として適していると考えられる。

４．本研究の目的と分析設計

4-1. 本研究の目的と意義

本研究の目的は、同社のサビキ需要を対象に、Prophet の予測結果に生成 AI による定性的な解釈・補正提案を組み合わせた手法を構築し、その効果を Prophet 単体および実際の実績データとの比較によって定量的に検証することである。本研究の意義は、技術的な新規性ではなく実務における検証可能性にあり、電力需要予測の知見をレジャー依存的なサビキ需要へ適用する可能性を検証するとともに、Prophet と生成 AI を代替ではなく補完的な役割分担として位置づけ効果を定量検証する点にある。ただし、本研究は単一企業・単一製品群を対象とした事例研究であり、一般化可能性の検討は今後の課題として残される。

4-2. リサーチクエスチョン

本研究では、次の２つのリサーチクエスチョンに答える。生成 AI による補完を組み

合わせた需要予測は、モデルベース型時系列予測手法と比較してどのような需要特性や外部環境条件の下で有効に機能するか（RQ１）、また、生成 AI を活用した需要予測は企業の在庫戦略や販売計画における意思決定支援にどのように貢献しうるか（RQ２）、である。

RQ１については、Prophet 単体による予測結果と、Prophet に生成 AI による補完を加えた予測結果とを、後述するローリング型バックテストにより比較し、「生成 AI による補完を加えた需要予測は、Prophet 単体による予測と比較して、より高い予測精度を示す」という仮説（H１）を設定する。なお、RQ１における「有効に機能する」とは、本研究では仮説 H１の通り、Prophet 単体と比較した予測精度、具体的には WAPE によって測定される誤差率の低さを指す。解釈可能性・納得性・迅速性といった精度以外の観点については、RQ２に関する考察の中で扱う。

RQ２については、現場担当者へのヒアリング等の定量的な検証手段を確立していないため、仮説検証としてではなく、本研究の分析結果に基づく考察の視点として位置づける。

4-3. 説明変数の選定根拠

説明変数は、時系列要因（トレンド・季節性・祝日等）、外部環境要因（気温・降水量・為替・輸入物価指数等）、行動要因（休日・連休・利用シーン）、商品属性要因（針サイズ・ハリス仕様・用途等）、補助的イベント要因（コロナ特需・規制変更等）の５つに整理する。確定的なカレンダー情報である時系列要因は Prophet に直接投入し、天候・行動条件・イベント情報のように解釈を要する定性的要因は生成 AI に文脈情報として与える。商品属性要因は、SKU 単位の分析単位を定める前提情報として両者に共通して用いる。気温・降水量・為替・輸入物価指数は、季節予報等の予測時点の見通しではなく気象庁・日本銀行の確定値を代理変数として用いる簡略化を採用しており、この結果は生成 AI による補完効果を実際よりも高く見積もる方向に偏る可能性がある。この点は 6-4 節で改めて述べる。

4-4. 本研究における分析設計と評価指標

分析対象は、同社が継続的に販売する国内向けサビキ製品群であり、SKU 単位の月次受注数量を予測単位とする。比較対象は、Prophet 単体による予測と、Prophet の予測値に生成 AI による補正を加えた予測の２種類である。生成 AI には Claude Sonnet 4.6 を用い、パラメータは API デフォルトとした。SKU 属性・Prophet 予測

値・外部環境要因・行動要因・補助的イベント要因を提示し、補正の方向（増加・減少・維持）・割合・根拠・確信度を JSON 形式で出力させ、あらかじめ定めたルールに従って予測値へ機械的に適用する。具体的な出力仕様・適用手順は 5-3-1 節で述べる。評価は、2019 年～2025 年の 7 年分の受注データを用い、学習データを 1 年ずつ拡張しながら 2023 年・2024 年・2025 年を予測対象とするローリング方式のバックテストにより行い、評価指標には WAPE[2]を中心に用いて、Prophet 単体と Prophet＋生成 AI 補完の誤差を SKU 単位・製品群全体の両面から比較する。また、受注データと販売実績データの差分から欠品による機会損失を分析するとともに、確信度が「低」の提案を一律に除外した場合の予測精度への影響を感度分析として確認する。なお、Prophet の予測値が理論上負の値を取り得る点への対応として、出力段階で 0 未満の値を 0 にクリップした上で生成 AI への入力に用いており、このクリップ処理自体が Prophet 単体の精度にも影響するため、5-2 節と 5-3 節以降とでは比較の基準となる Prophet 予測値が異なる点に留意されたい。この詳細は 5-3-2 節で述べる。

本研究の分析設計は、渡邉ほか（2020）や中山・浦野（2023）が採用したローリング方式のバックテストという評価枠組みを踏襲する一方、これらの先行研究が主に時系列モデル単体の精度検証にとどまっていたのに対し、本研究では Prophet の予測値に生成 AI による定性的な補正を組み合わせ、その適用方式（乗算方式・加法的フォールバック）自体を比較対象に含める点で独自性を持つ。この独自の設計により、生成 AI の補完効果を、判断能力そのものの効果と、その判断を予測値へ反映する実装上の設計の効果とに分けて検証することができる。

なお、本研究における一連のデータ処理、Prophet による予測、評価指標の算出は、プログラミング言語 Python（pandas、Prophet ライブラリ等を使用）を用いて実装した。

５．需要予測手法の適用と比較分析

5-1. 分析データの概要

分析対象は、4-4 節で述べた通り、A 社が継続的に販売しているサビキ製品群のうち、国内 47 都道府県への納品分である。対象期間は 2019 年 1 月から 2025 年 12 月までの 7 年間であり、月次に集計した受注数量を分析の基本単位とする。

[2] WAPE(Weighted Absolute Percentage Error、加重絶対誤差率):月ごとの予測誤差(絶対値)をすべて合計し、それを実績値の合計で割って百分率にした指標。実績の規模が異なる SKU 間でも公平に精度を比較できる。

なお、対象企業の機密性および匿名性を確保する観点から、分析対象とするSKUの正確な数や、各SKUのデータ点数等の規模を示す詳細な数値は、本論文中には記載しない。ただし、分析対象の規模感を示すため付言すると、対象となるサビキのSKUは数百点規模である。分析自体は実際の受注データを用いて行うが、論文中で示す結果は、予測誤差や精度指標、改善率等に加工した形でのみ提示する。

また、対象期間中にはSKUの新設・廃止が生じており、SKUによって取得可能なデータ点数が異なる。受注数量が観測されない月、すなわち受注ゼロの月も一定程度存在しており、こうした特性は、5-2節以降で示す予測精度の評価に影響を及ぼす可能性があるため、解釈にあたって留意する必要がある。

5-2. Prophetによる需要予測

5-2-1. 予測結果の概要

Prophetの設定にあたっては、トレンドの柔軟性を表すパラメータであるchangepoint prior scale、年次季節性の有無、祝日効果の有無を組み合わせた8パターンについて、2023年・2024年の2時点、すなわち決定用データにおける平均WAPEを比較する感度分析を行った。感度分析の結果を表2に示す。

表2：8パターンの設定とバックテスト結果

パターン	トレンド設定	年次季節性	祝日効果	第1回WAPE	第2回WAPE	第3回WAPE
A	標準	なし	なし	127.1%	49.7%	47.2%
B	標準	あり	なし	125.9%	66.0%	49.8%
C	標準	なし	あり	126.3%	48.0%	45.0%
D	標準	あり	あり	124.2%	65.8%	49.6%
E（採用）	**柔軟**	**なし**	**なし**	**79.0%**	**79.8%**	**50.5%**
F	柔軟	あり	なし	105.7%	92.8%	53.4%
G	柔軟	なし	あり	78.7%	80.6%	50.1%
H	柔軟	あり	あり	108.3%	91.7%	53.5%

標準的な設定としてProphetの初期値0.05を、柔軟性を高めた設定としてその10倍の0.5を用いた。標準的な設定では、学習期間全体を通じてなだらかに変化する一つのトレンドとして需要を捉えるのに対し、柔軟性を高めた設定では、学習期間の中

でも直近の局所的な増減により強く反応し、トレンドの折れ線を細かく引き直しながら需要を捉える。

感度分析の結果、最も平均 WAPE が低かったのは、トレンドの柔軟性を高め、年次季節性および祝日効果を含まない設定、すなわちパターン E であった。この設定を採用パターンとして決定し、未知のデータである 2025 年、すなわち検証用データに適用した結果、決定用データでの平均 WAPE と検証用データでの WAPE の間に大きな差は見られず、特定の年に過剰適合した設定ではないことが確認された。この結果を図２に示す。実際、決定用データにおける WAPE の標準偏差を確認すると、採用パターン E は 0.6 ポイントと極めて小さいのに対し、標準的なトレンド設定である A〜D 群は 41〜55 ポイントに達しており、標準設定は学習期間の取り方によって精度が大きく左右されやすいことを示している。

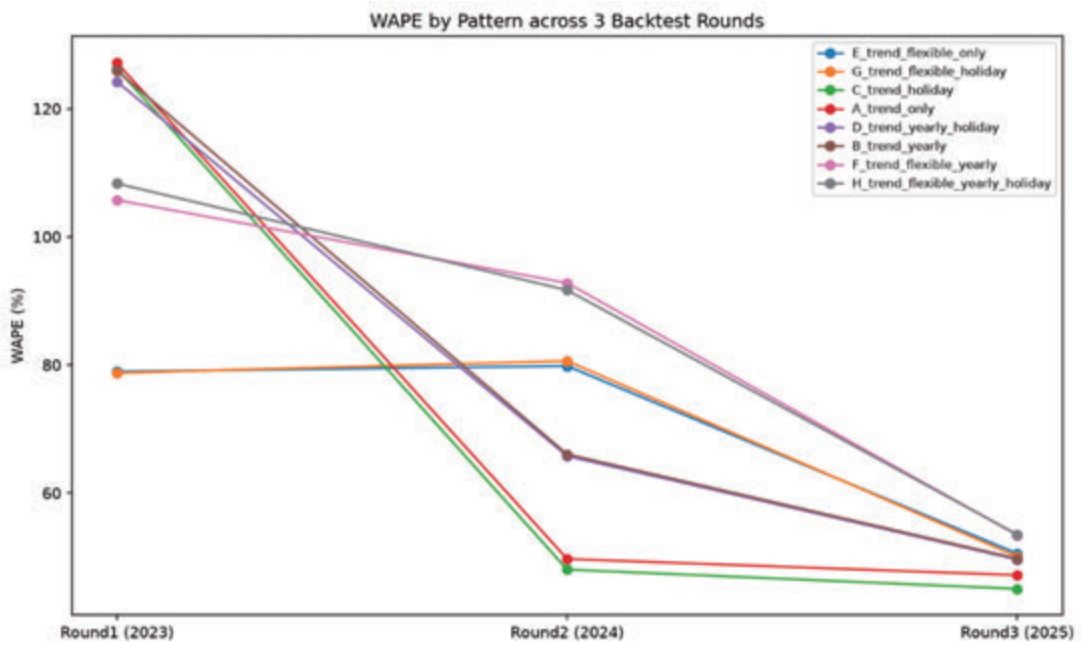

図２：８パターン×３時点における WAPE の推移

（出所：Python にて筆者出力）

採用した設定による３回のバックテストの全体精度を確認する。予測対象は第１回が 2023 年、第２回が 2024 年、第３回が 2025 年であり、全体 WAPE はそれぞれ 79.0%、79.8%、50.5%であった。学習データの期間が最も長い第３回で精度が大きく改善し

ており、学習データの蓄積量が予測精度に影響を与える可能性が示唆される。この点は、データ量が少ない場合に精度が低下しやすいという先行研究の指摘（弓岡，2023）と整合する。

5-2-2. 時系列構造の可視化と解釈

図２に示した８パターン×３時点の WAPE の推移を見ると、８パターンは、トレンドの柔軟性を高く設定したグループ E〜H と、標準的なトレンド設定によるグループ A〜D に大別される。標準的なトレンド設定は、第１回で WAPE 約 124〜127％という著しく低い精度を示す一方、第２回・第３回では約 45〜66％まで改善するなど、回によって精度が大きく変動した。これに対し、採用パターン E および G は、３回を通じて約 50〜81％という比較的安定した精度を示している。

この違いは、対象期間における受注数量全体の推移と関連していると考えられる。3-2 節で述べた通り、同社の釣具需要は、新型コロナウイルス感染症拡大期における屋外レジャー需要の高まりとその後の反動減という、平常時とは異なる局面を経験しており、年間受注数量は 2020 年・2021 年に前年比約＋49％、約＋15％と大きく増加した後、2022 年・2023 年にかけて約－31％、約－40％と急減し、2024 年に約＋７％とやや回復した。第１回のバックテストは、この急増から急減への転換点を含む期間を学習データとしており、需要変化に鈍感な標準設定ではこの転換を十分に捉えられず、2023 年の実績を大きく見誤ったと考えられる。一方、柔軟なトレンド設定は、学習期間内の直近の下降局面に敏感に反応するため、2023 年の継続的な減少をある程度捉えることができた。

ただし、この柔軟性が常に有利に働くわけではない。第２回のバックテストでは、学習データが 2023 年の下降局面の末端まで及ぶため、柔軟な設定はこの下降傾向をそのまま 2024 年まで延長して予測する形となった。しかし実際の 2024 年の受注数量は回復に転じており、その結果、多くの SKU で予測値が実績を大きく下回る系統的な過小予測が生じている。第３回では、学習データが 2024 年の回復局面まで含むため、いずれの設定でも比較的安定した精度に収束している。

加えて、年次季節性を加えたパターン F・H は、対応する季節性なしのパターン E・G よりも、３回すべてで一貫して WAPE が高く精度が低いことも確認された。これは、「春から秋にかけて需要が高まり、冬季に低下する」というサビキ需要の季節性が、年によって安定的に繰り返される滑らかな周期構造としては現れにくく、天候や海況

等による年ごとの不規則な変動の影響を強く受けていることを示唆している。祝日効果の有無による差は、３回を通じていずれも１ポイント未満と限定的であった。

以上より、サビキ需要の SKU 単位の月次受注データに対しては、固定的な季節性を明示的に組み込むよりも、トレンドの局所的な変化を柔軟に捉えることの方が、予測精度および頑健性の観点から重要であるといえる。これは、コロナ特需のような特殊局面を時系列構造の一部として単純に吸収できるとは限らないという Prophet の限界とも整合する結果であり、Prophet 単体による時系列構造の把握には、なお一定の限界が残ることを示している。

5-2-3. SKU 別の特徴と課題

全体 WAPE は受注数量で加重した平均であるため、SKU 単位の予測の振れ幅は表れない。SKU 単位の予測/実績比の分布を確認すると、全体 WAPE が近い水準であった第１回と第２回の間で、内訳には大きな違いが見られる。第１回では予測/実績比の中央値は約 86％であったのに対し、第２回では、実績を過小に予測した SKU が全体の９割超を占め、中央値も約 22％まで低下した。これは、2022 年から 2023 年にかけての急減局面を学習データの末端に含む状態で 2024 年を予測したため、多くの SKU で下降トレンドが過度に延長され、実際の需要回復を捉えられなかったことによる。すなわち、第２回の全体 WAPE が第１回と近い水準だったのは、受注量の大きい一部 SKU の予測が相対的に良好だったためであり、SKU 単位で見れば、予測の系統的な偏りはむしろ第２回で最も顕著であった。

製品カテゴリー別では、船用サビキが堤防・岸釣り用サビキよりも全体的に WAPE が高く精度が低い傾向が確認された。この傾向を図３・図４に示す。SKU 単位の WAPE 中央値は、堤防・岸釣り用が第１回・第２回・第３回でそれぞれ約 77％、約 83％、約 53％であるのに対し、船用は約 92％、約 101％、約 71％と、いずれの回でも堤防・岸釣り用を上回る。この差は、波が高く船が出航できない場合には釣行自体の中止につながるという海況の影響と関連していると考えられる。また、堤防・岸釣りは天候・海況が多少悪化しても釣行が継続可能な場合が多いのに対し、船釣りは出航の可否という明確な閾値的条件によって釣行の有無が決まりやすく、より非連続的な需要構造を持つため、連続的な時系列構造を前提とする Prophet では捉えにくいと考えられる。

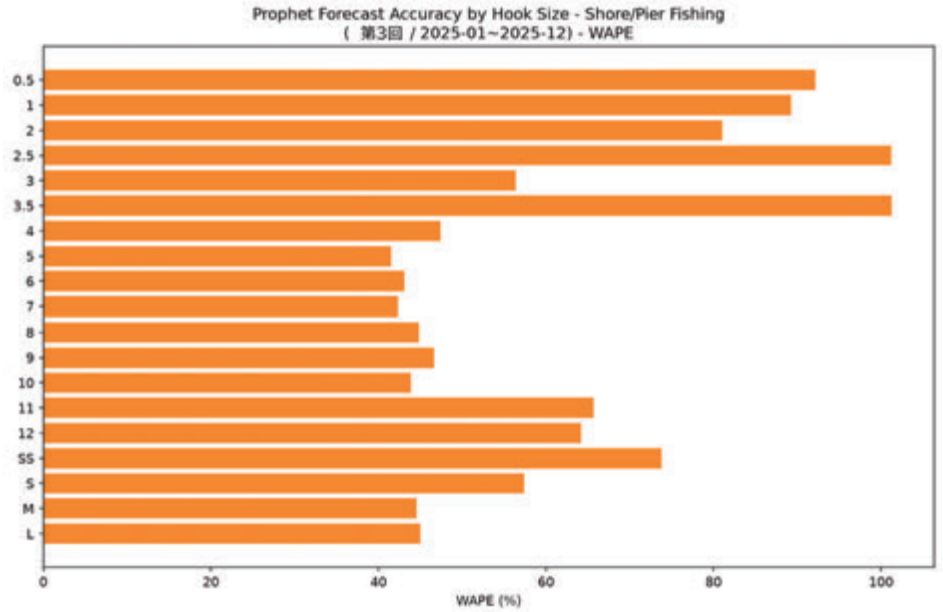

図３：堤防・岸釣り用サビキ針サイズ別 WAPE（2025 年予測）

（出所：Python にて筆者出力）

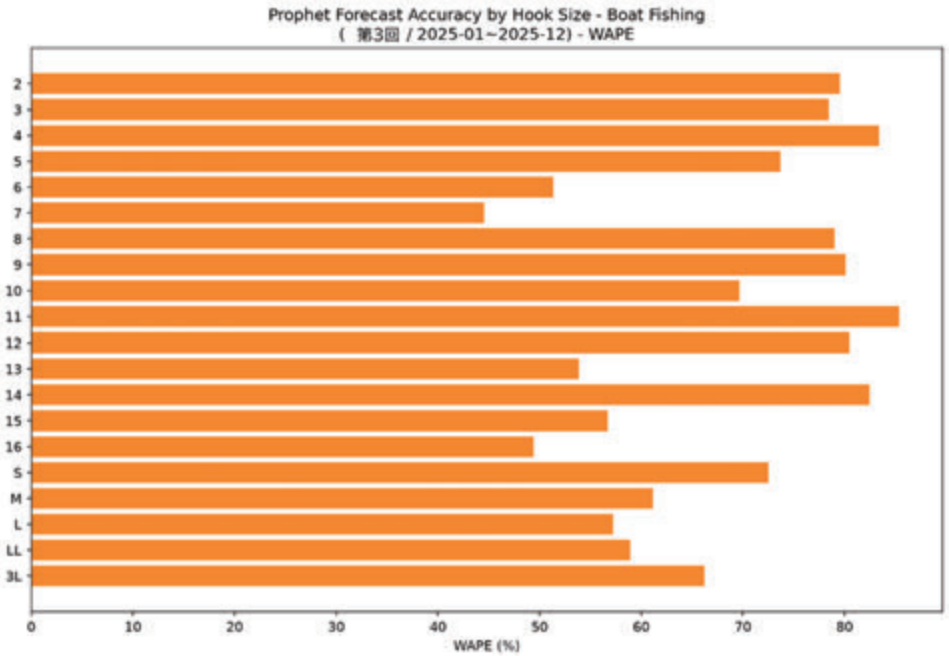

図４：船用サビキ針サイズ別 WAPE（2025 年予測）

（出所：Python にて筆者出力）

以上から、サビキ需要は製品全体としては一定の時系列構造を有するものの、SKU単位、とりわけ船用サビキにおいてはProphet単体による予測の限界が顕著に表れることが明らかとなった。次節では、こうした課題に対し、生成AIによる定性的な解釈・補正がどのように補完的に機能するかを検討する。

5-3. 生成AIによる補完後の需要予測

生成AIによる定性的な解釈・補正を加えた予測結果を、以下「Prophet＋生成AI補完」と呼ぶ。4-4節で述べた通り、生成AIには、Prophetの予測値に加えて、外部環境要因、行動要因、商品属性要因、補助的イベント要因を定性情報として与え、補正の方向性・割合・根拠・確信度を構造化された形式で出力させた上で、あらかじめ定めた計算ルールに従って予測値に反映させた。この補正は、5-2節と同一のローリング方式のバックテストである2023年・2024年・2025年の３時点に対して、同一の学習期間・予測対象期間・SKU単位データを用いて実施した。

5-3-1. 生成AIへの入力情報と補正プロセス

生成AIには、SKU単位・月単位で、SKUの属性情報である品番、製品カテゴリー、針サイズ、ハリスサイズ、商品ランク、シリーズ、Prophetによる当該月の予測値、予測対象期間に先立って判明している外部環境要因である納品先都道府県別の平均気温・降水量、為替レート、輸入物価指数、当該月の祝日数・休日数・最大連休日数といった行動要因、判明している範囲での補助的イベント情報を与える。これらの情報は、4-4節で述べた通り、各SKU・各月で同一の構成・順序により提示し、生成AIには、補正の方向である増加・減少・維持、およびおおよその割合を０～100％の範囲で、判断の根拠および確信度である高・中・低とともに、構造化された形式であるJSON形式で出力させる。具体的なプロンプトは付録に示す。

このプロンプト設計の妥当性は、2-2節で述べた本研究の比較設計上の要請という枠組みから説明できる。すなわち、生成AIには補正の方向性と割合の提案のみを担わせ、最終的な数値計算は筆者があらかじめ定めた固定ルールに委ねている。これは、Prophetを解釈可能性の高さゆえに比較の基準として採用したのと同じ考え方に基づく。具体的には、生成AIの判断の質による効果と、適用方式という計算ルールの設計による効果とを、切り分けて検証するためである。仮に生成AIに数値計算そのものまで委ねた場合、精度の変化が生成AIの判断の質によるものか、その計算過程によるものかを切り分けられなくなり、後述する5-4-1節で示すように、判断能力

の効果と適用方式の効果を分離する分析、すなわち後述の表３で示す乗算方式と加法的フォールバックの比較が成立しなくなる。また、判断根拠に対する確信度（高・中・低）を出力させ、根拠が乏しい場合には確信度を「低」とするよう指示している点も、5-4-1 節で示す確信度別の精度検証という感度分析を可能にするための設計であり、同様に比較・検証を目的としたものである。なお、与えられた情報以外の事後情報を用いないよう明示的に指示している点は、2-3 節で述べた通り、生成 AI の出力が必ずしも事実に基づいた正確なものとは限らないというハルシネーションのリスクを踏まえた対策である。

出力された補正の方向性と割合は、4-4 節で述べた手順に従って、Prophet の予測値に機械的に適用する。方向が「増加」の場合は予測値に（１＋pct／100）を、「減少」の場合は（１－pct／100）を乗じ、「維持」の場合は補正を行わず Prophet の予測値をそのまま採用する。この一連の手順は、Claude 上に構築した専用アプリケーションを通じて SKU・月単位で反復実行される。

5-3-2. 補完後の予測結果の概要

生成 AI への入力には、Prophet の出力段階で０未満の値を０にクリップした予測値を用いている。このクリップ後の Prophet 予測の WAPE は、第１回 69.9%・第２回 68.9%・第３回 49.3%であり、5-2 節で示した未クリップの数値を下回る。このクリップ後の予測を基準として、生成 AI の補正提案を機械的に適用したところ、乗算方式のみでは、第１回は WAPE が 71.1%とかえって悪化し、第２回は 67.8%、第３回は 48.8%と、わずかな改善にとどまった。この結果を表３に示す。

表３：３ラウンドの全体精度比較（WAPE）

	Prophet（クリップ後）	乗算方式のみ	加法フォールバック追加
第１回	69.9%	71.1%　（－1.2pt）	67.6%　（＋2.3pt）
第２回	68.9%	67.8%　（＋1.1pt）	62.7%　（＋6.2pt）
第３回	49.3%	48.8%　（＋0.5pt）	48.1%　（＋1.2pt）

この背景には、Prophet の予測値が 0 となる SKU・月では、生成 AI が増加・減少といった方向転換を提案しても、（1 ＋pct／100）または（1 －pct／100）を 0 に乗じる限り値が変化しないという、乗算方式固有の構造的な限界がある。表 4 に示す通り、ゼロ予測となった SKU・月の約 7 割で生成 AI は増加を提案しており、これらは乗算方式では反映されない。ゼロ予測率自体も第 2 回に 34.0%まで急増しており、この局面で乗算方式の限界が最も顕在化した。ゼロ予測率が第 2 回に急増しているのは、学習データが 2023 年の急減局面の末端まで及び、下降トレンドが 2024 年まで過度に延長されたためである。

表 4 ：ゼロ予測の発生状況と生成 AI 補正提案の内訳

	ゼロ予測率	方向転換提案（対象）			対象外
		増加	減少	小計	維持
第 1 回	15.7%	74.9%	0.2%	75.1%	24.9%
第 2 回	34.0%	71.2%	0.1%	71.3%	28.7%
第 3 回	9.5%	67.5%	0.0%	67.5%	32.5%

そこで、Prophet の予測値が 0 となる場合に限り、当該 SKU の学習期間中の月次受注平均値に補正割合を乗じて予測値とする加法的フォールバックを追加した。この改良は、生成 AI が既に出力した補正提案を再利用するものであり、追加の API 呼び出しを要しない。なお、このルールは 3 回すべての乗算方式適用結果を確認した上で設計しており、検証対象期間全体の情報を参照しているため、後述する「3 回すべてで上回った」という結果を、未知データへの厳密な事前検証と同一視することはできない点に留意されたい。この改良により、3 回すべてで Prophet 単体を上回る結果である 67.6%、62.7%、48.1%が得られたことは、表 3 の通りである。

また、表 5 の通り、SKU 単位でも、改善した SKU の割合が悪化した SKU の割合を 3 回すべてで上回った。符号検定、すなわち二項検定・片側検定でも 3 回すべてで統計的に有意な差が確認された。これは、全体 WAPE の改善が一部の大口 SKU のみによるものではなく、SKU 単位でも広く観察される傾向であることを示している。

表５：SKU 単位の改善率と符号検定結果

	改善 SKU 割合	悪化 SKU 割合	有意水準（片側二項検定）
第１回	55.6%	44.4%	p=0.008
第２回	87.5%	12.5%	p<0.001
第３回	76.1%	23.7%	p<0.001

注：第３回は改善・悪化が同水準の SKU が１件含まれるため、合計は 100%に満たない

5-3-3. SKU 別の特徴と課題

表６は、加法フォールバック追加後のカテゴリー別 WAPE 推移を示す。堤防・岸釣り用は３回とも改善し、改善幅は第１回 3.0 ポイント、第２回 5.5 ポイント、第３回 0.6 ポイントであった。船用は第２回・第３回で大きく改善し、それぞれ 10.1 ポイント、4.4 ポイントの改善幅を示した一方、第１回はわずかに 0.8 ポイント悪化した。船用サビキは出航可否という閾値的条件で需要が非連続に変動するため、学習期間平均を基準値とする加法方式でも変動を捉えきれない場合がある。それでも、ゼロ予測率が低下した第２回・第３回で改善が顕著であり、Prophet のゼロ予測というボトルネックが解消されて初めて生成 AI の補完効果が発揮される構造にあるといえる。

表６：カテゴリー別 WAPE 推移（クリップ後 Prophet→生成 AI 補正後、改良版）

	堤防･岸釣り用	船用
第１回	66.7% → 63.7%（＋3.0pt）	85.2% → 85.9%（－0.8pt）
第２回	67.3% → 61.8%（＋5.5pt）	77.9% → 67.8%（＋10.1pt）
第３回	46.0% → 45.5%（＋0.6pt）	66.1% → 61.7%（＋4.4pt）

なお、全 SKU・全３回分の補正提案取得には自作のバッチ処理アプリケーション（genai_correction_runner.html）を用いたが、Anthropic API のレートリミットにより処理が自動停止と手動再開を繰り返し、想定より大幅に多くの作業時間を要した。この点は、6-4 節で改めて述べる。

5-4. Prophet 単体と生成 AI 補完後の比較

5-4-1. 予測精度の比較

表３の通り、改良後の生成 AI 補完は３回すべてで Prophet 単体を上回り、改善幅は第１回 2.3 ポイント、第２回 6.2 ポイント、第３回 1.2 ポイントであった。ただし、事前に設計した乗算方式のみでは第１回はむしろ悪化しており、３回すべてで上回ったのは、検証データの結果を参照して事後的に設計した加法的フォールバックによるものである。この点で、本結果を仮説 H１の厳密な検証結果と同一視することはできない。精度改善の成否を分けたのは生成 AI の出力の質そのものよりも、その出力を予測値へ適用する方式であったといえる。改良効果はゼロ予測率が 34.0%と最も高かった第２回で最大の 6.2 ポイント改善となり、Prophet 単体の精度が低下しやすい局面ほど補完の効果が大きいことが示された。

表６の通り、カテゴリー別でも、需要が非連続的な船用サビキにおいて、ゼロ予測率が低下した第２回・第３回に 10.1 ポイント、4.4 ポイントという大きな改善が確認された。また、確信度「低」の提案を除外すると全回で精度が向上し、改善幅は第１回 4.7 ポイント、第２回 6.6 ポイント、第３回 1.5 ポイントであったものの、除外後も実務導入水準には遠く、確信度は精度向上の手段ではなく優先順位づけの手がかりにとどまる。ただし、この確信度別の除外効果は、３回の結果を確認した上での事後的な確認であり、事前に確立された除外基準に基づくものではない点に留意が必要である。この点は、加法的フォールバックと同様の限界を伴う。

生成 AI による補正の方向は、表７の通り、全体で増加が 59.1%と過半を占め、維持 22.8%、減少 18.1%の順であった。

表７：AI 補完の方向別割合

	増加	維持	減少
第１回	61.6%	24.0%	14.3%
第２回	60.0%	25.7%	14.3%
第３回	55.5%	18.7%	25.7%
全体	59.1%	22.8%	18.1%

一方、SKU・月単位では、増加または減少へ方向を変更した提案のうち 34.0%で Prophet 単体より誤差が悪化しており、表８の通り、Prophet の予測値がすでに実績

を上回っている場面での増加補正は対実績比 100〜130%で 99.9%、130%を超える場合は 100%のケースで誤差を悪化させ、実績を下回っている場面での減少補正は、対実績比 100%以下のいずれの区分でも全てのケースで誤差を悪化させた。これは、生成 AI が Prophet の予測値を入力として受け取ってはいるものの、その妥当性を補正の判断に反映できていないことに起因する。なお、表８の各区分は、いずれも数百件から数千件規模の SKU・月を含んでおり、0.0%や 100.0%といった値は、サンプル数が少ないことによる偶然ではなく、当該区分において補正の効果がほぼ一方向に定まることを示している。また、この 34.0%という悪化率は、5-3-2 節・表５で示した SKU 単位の改善率、すなわち改善した SKU の割合が第１回 55.6%、第２回 87.5%、第３回 76.1%であったという結果と一見矛盾するようにみえるが、集計単位が異なる点に注意が必要である。5-3-2 節の指標は SKU 単位で３回のテスト全体をロールアップした改善・悪化であり、同一 SKU 内で月によって悪化と改善が混在していても、年間を通じて改善が優勢であればその SKU は「改善」に分類される。一方、34.0%は SKU・月という最も細かい粒度で悪化の有無を個別に集計したものであり、月次の悪化が年間の SKU 単位集計では相殺されうるため、両者は矛盾しない。

表８：Prophet より誤差が悪化した割合（対実績比別）

Prophet 予測の対実績比	増加の悪化率	減少の悪化率
50%以下	0.0%	100.0%
50%超 70%以下	0.0%	100.0%
70%超 100%以下	22.1%	100.0%
100%超 130%以下	99.9%	31.1%
130%超	100.0%	0.0%

5-4-2. 解釈可能性の比較

Prophet は数値の内訳（トレンド・季節性・休日効果）は示すが、その背景要因までは説明しない。生成 AI は根拠（reason）を自然言語で提示し、例えば「冬季低温で需要低迷」のように、担当者が専門的な分析スキルを要さずとも背景を直感的に理解しやすい。

表９の通り、同一商品・同一月である５月・GW 期の全３回分で、Prophet 予測の対実績比が 320%→89%→69%と大きく変動する中でも、生成 AI は「GW 連休による需要ピーク」という一貫した判断軸を保ちつつ、20%、30%、20%と補正割合を調整しており、担当者の経験的判断を明示的に再現している。ただし、割合の算出根拠自体は説明されず、これは大規模言語モデルに共通する限界である。生成 AI の価値は補正割合の精緻さではなく、天候・イベント等の定性的要因を判断の俎上に載せる点にあり、割合の妥当性は担当者による事後確認を要する。

表９：生成 AI 補正の具体例

		AI の根拠（reason）	補正方向・割合（確信度）	Prophet 予測の対実績比	補正後予測の対実績比	加法フォールバック後
商品 A	第１回	GW 連休・釣行需要ピーク	増加・20%（中）	320%	384%	—
	第２回	GW 祝日４日·気温好適で需要ピーク	増加・30%（高）	89%	116%	—
	第３回	GW４連休で家族向け需要増加見込み	増加・20%（中）	69%	83%	—
商品 B	第１回	GW 連休・需要ピーク期	増加・20%（中）	85%	101%	—
	第２回	GW 連休･気温適期、０予測を修正	増加・30%（中）	0%	0%	34%
	第３回	GW４連休・釣りシーズン本格化	増加・20%（高）	60%	72%	—

5-4-3. 納得性・迅速性の比較

Prophet の数値は担当者が自ら解釈する必要があり、現行の実務と大差ない。生成 AI の根拠は担当者が普段用いる判断材料と重なりやすく、納得感を得やすいと考えられるが、これはヒアリング等で直接測定したものではない。迅速性については、処理自体は数秒で完了するが、API のレートリミットにより中断・再開を繰り返す必要が生じ、想定より大幅に多くの作業時間を要した。

5-4-4. 意思決定支援上の有効性に関する考察

改良後もなお全体 WAPE は 67.6%・62.7%・48.1%であり、実務への即時導入を無条件に提案できる水準には達していない。これは A 社の BtoB 構造による最終需要の

間接的把握という、Prophet・生成 AI 共通の制約に起因する部分が大きい。一方、生成 AI が示す根拠は、Prophet の数値のみでは伝わりにくい背景を自然言語で示す点で、解釈可能性・納得性を高める可能性があり、担当者による確認・優先順位づけを補助しうると考えられる。ただし、これはヒアリング等による現場担当者への検証を経た結果ではなく、あくまで本研究の分析結果に基づく考察である。したがって、生成 AI の価値は精度を実務水準まで引き上げる主役としてではなく、確信度が低い SKU やゼロ予測となった SKU を担当者が優先的に見直す候補として示す、二次的なチェック・説明ツールとして位置づける方が実態に即している。ただし、全 SKU への適用には API 利用制限が障害となりうる。以上の評価は単一モデルによる単発の出力に基づくものであり、実務での運用にあたっては、出力の安定性や他モデルでの再現性を確認した上で導入の可否を判断する必要がある。

なお、欠品疑いフラグ別の WAPE 比較について、表 10 に示す通り、サンプルの十分性にばらつきがあり傾向も一致せず、統計的に頑健な結論は得られなかった。

表 10：欠品疑いフラグ別の WAPE 比較

	欠品疑いなし	欠品疑いあり	サンプルの十分性
第１回	70.40%	63.20%	●●●（十分）
第２回	69.00%	45.70%	●○○（少ない）
第３回	49.30%	69.30%	○○○（極めて少ない）

5-5. 比較分析から得られる示唆

仮説 H１は、事後的に設計した適用方式を用いるという条件の下でのみ支持された。すなわち、事前に設計した乗算方式のみでは効果が不安定であったが、検証データの結果を参照して設計した加法的フォールバックにより、すべての回で Prophet 単体を上回った。ただし、この「事後設計」という条件は、厳密なアウトオブサンプル検証ではないという限界を伴うため、本研究は H１を暫定的に支持するにとどまる。むしろ、この結果が明確に示すのは、補完効果が生成 AI の判断能力そのものよりも、その出力を予測値へ適用する方式に強く依存するという点である。加えて、SKU・月単位では方向を変更した提案の 34.0%でかえって誤差が悪化しており、生成 AI の補正が Prophet の現状の予測水準を考慮しない場合には効果が反転しやすいこと

も明らかになった。RQ１には「Prophet 単体の精度が低下しやすい局面で、かつ適切な適用方式を伴う場合に有効に機能する可能性がある」と回答できる。

RQ２は、精度面で実務への即時導入水準には至っていないが、自然言語の根拠は解釈可能性・納得性を高める可能性があり、担当者による確認・優先順位づけを補助しうると考えられる、ただし全 SKU 適用には API 利用制限という制約が残る、と回答できる。

６．結論と考察

6-1. 本研究の結論

本研究は、A 社が抱える需要予測上の課題である、経験や勘への依存、予測作業に要する多大な時間、外生要因を反映しにくいという課題を出発点とし、Prophet による時系列予測に生成 AI による定性的な解釈・補正を組み合わせた手法を構築し、Prophet 単体との比較を通じてその効果を検証した。生成 AI 補完は、実務への即時導入を提案できる水準には至らなかったものの、適用方式を工夫することで、精度の面で Prophet 単体を上回りうることを示した。ただし、この上回りは検証データを参照して設計した適用方式によるものであり、厳密なアウトオブサンプル検証ではない点には留意を要する。加えて、担当者が経験や勘に基づいて行ってきた需要判断を、天候・海況・イベントといった根拠を伴う明示的な形式知として再現・可視化できた可能性も、本研究が示唆する点である。

ただし、この形式知化は常に有効に働くわけではなく、Prophet の現状の予測水準を考慮しない補正が、SKU・月単位で見れば３割超のケースでかえって精度を悪化させていたことも、本研究が明らかにした重要な留意点である。5-4-2 節で示した通り、生成 AI が出力した判断根拠は担当者の暗黙の判断過程と類似しており、属人化した判断プロセスを言語化する意思決定支援ツールとして機能しうることを示している。予測精度である WAPE の改善は、この形式知化が実際の需要変動とどの程度整合していたかを示す、副次的な検証結果として位置づけられる。

なお、以上の形式知化に関する解釈は、生成 AI が提示する判断根拠と、3-1 節で述べたインタビューにおける同責任者の発言との対応関係に基づく筆者の分析であり、関係者自身による検証を経たものではない。改良後もなお実務導入水準に達していない現状の精度で分析結果を関係者に開示し、評価を求めることは時期尚早と判断し、本研究では関係者からのフィードバックの取得は行っていない。この解釈の妥当性を

関係者自身の評価によって検証することは、6-5 節で述べる今後の課題としたい。

6-2. 学術的意義

本研究の学術的意義は、先行研究が抱えていた限界と、本研究がそれに対して付け加えた点を対応づけることで、次の３点に整理できる。

第一に、電力需要予測分野で蓄積されてきたモデルベース型時系列予測手法の知見は、需要が連続的・安定的に発生することを前提としており、非連続的かつレジャー依存的な需要構造への適用可能性は、これまで十分に検証されてこなかった。本研究は、この限界に対し、サビキ需要という具体的な非連続的需要構造を対象に実データで検証を行うことで、電力需要予測の知見が適用可能な範囲と、適用に際して生じる限界、すなわちゼロ予測の壁を具体的に明らかにした。これは、需要予測研究の対象を、安定的な需要を持つ財から、より不確実性の高い財へと拡張する一事例を提供するものであり、この知見は、レジャー依存的な需要を持つ他の製品やサービスへの適用可能性を検証する研究へとつながるものである。

第二に、Prophet と生成 AI は、それぞれの特徴と限界が個別に検討されてきた一方、両者を補完的な役割分担として組み合わせ、その効果を定量的に検証した研究は乏しかった。本研究は、この限界に対し、乗算方式とその限界に対応した加法的フォールバックという具体的な適用方式を設計し、この設計の違いが精度に与える影響を、SKU 単位の統計的検定を含めて定量的に明らかにした。これは、生成 AI を需要予測に組み込む際の実装上の設計が、生成 AI そのものの性能と同等以上に精度を左右しうるという知見を提供するものであり、今後、他の適用方式や、他の基準モデルを用いた場合の効果検証という課題を示唆している。

第三に、需要予測研究は、予測精度の改善を主な評価軸とすることが多く、意思決定支援としての価値という観点は軽視されがちであった。本研究は、この限界に対し、解釈可能性・納得性・迅速性という多面的な観点から、生成 AI の実装経験に基づいて具体的に記述した。これは、精度指標だけでは捉えられない、生成 AI 活用の実務的な論点を需要予測研究の文脈に提供するものであり、今後、現場担当者へのヒアリング等を通じて、本研究が考察にとどめた意思決定支援上の効果を実証的に検証する必要性を示すものである。ただし、以上の貢献の射程は限定的である点を明記しておく必要がある。本研究が定量的に示したのは、あくまで Prophet を基準モデルとし、生成 AI を補正手段として組み合わせるという特定の構成において、適用方式の設計が精度を左右するという知見である。間欠需要予測に特化した手法や他の基準モデルを用いた

場合にも同様の知見が成立するかは、本研究の範囲外であり、6-5 節で述べる今後の課題に属する。本研究の貢献は、この限定された構成の下で、適用方式という従来見過ごされてきた設計要素の重要性を、実データによる統計的検証を伴って具体的に示した点にある。

これらの今後の研究の方向性は、6-5 節で改めて具体的な課題として整理する。

6-3. 実務的示唆

本研究の結果を踏まえ、A 社をはじめとする同種の課題を抱える中小製造業に対し、次の実務的示唆を提示する。生成 AI を需要予測の自動化ツールとして無条件に導入することには慎重であるべきであり、得られた精度は実務上十分とは言えないため、担当者による確認を前提とした運用が現実的である。加えて、生成 AI の出力は非決定的であり、同一の入力でも実行のたびに提案が変動しうるため、実務で用いる際には、複数回の実行による出力の安定性を確認するか、変動しうることを前提とした運用設計が求められる。むしろ生成 AI は、確信度が低い SKU やゼロ予測となった SKU を担当者が優先的に見直すべき候補として提示する、チェック・説明ツールとして位置づけることで現実的な価値を発揮しうる。また、補正提案を予測値へ適用する計算方式の設計は、生成 AI 自体の性能と同等以上に重要であり、同一の出力でも適用方式を変えるだけで精度の優劣が逆転した点は特に留意すべきである。さらに、SKU 単位で大量運用する場合には、API 利用プランやレートリミットへの対応をあらかじめ整備しておく必要がある。

6-4. 本研究の限界

本研究には以下の限界がある。

第一に、本研究は A 社１社・サビキという単一製品群を対象とした事例研究であり、知見の一般化可能性は本研究の範囲を超える。

第二に、A 社は BtoB 取引を基本とし最終消費者の購買行動を流通事業者経由でしか把握できないため、需要予測の精度には構造的な上限が存在すると考えられる。

第三に、外部環境要因には事後確定値を代理変数として用いており、これは生成 AI による補完効果を実際よりも高く見積もる方向に偏る可能性がある。また、加法的フォールバックのルールは乗算方式の限界への応急的な対処であるとともに、３回すべての結果を確認した上で設計されており、検証対象期間全体の情報を参照しているため、「３回すべてで上回った」という結果は、未知データに対する厳密な意味でのア

ウトオブサンプル検証ではない。比較対象もProphetのみであり、XGBoost等のより高精度な手法との比較は行っていない。また、5-4-1節で確認した通り、生成AIの補正は方向を変更した提案の34.0%でかえって誤差を悪化させる副作用を伴う。

第四に、API利用制限により想定を大幅に上回る作業時間を要した点、生成AIの補正割合の算出過程が説明できないブラックボックスである点に加え、納得性・迅速性・意思決定支援上の有効性は現場担当者へのヒアリング等を経ない考察にとどまり、欠品機会損失の分析も統計的に頑健な結論には至らなかった。また、6-1節で述べた通り、生成AIの判断根拠が担当者の暗黙知を可視化したものであるという本研究の解釈自体も、関係者による検証を経ておらず、筆者の分析にとどまる点に留意が必要である。

第五に、生成AIの判断能力そのものに起因する限界である。5-4-1節で確認した通り、生成AIによる補正のうち、方向を変更した提案の34.0%は、Prophet単体よりもかえって誤差を悪化させていた。この限界は、方法論上の簡略化という第三の限界とは性質が異なり、生成AIの判断プロセスそのものの限界である。

第六に、本研究の知見は、単一の生成AIモデル（Claude Sonnet 4.6）による単発実行に基づくものである。生成AIの出力は非決定的であり、同一の入力に対しても実行のたびに出力が変動しうるが、本研究ではSKU・月ごとに一度のみ補正提案を取得しており、出力の安定性・再現性は検証していない。また、他の生成AIモデルを用いた場合に同様の結果が得られるかも検証していない。したがって、本研究が示した生成AIの判断に関する知見を、生成AIによる需要予測支援一般へと拡張するには、複数回の試行による出力の安定性の確認と、複数モデルによる再現性の検証が必要である。

6-5. 今後の課題

一般化可能性という第一の限界に対しては、他の企業・製品群への適用可能性を検証することが課題となる。特に、本研究で確認された「Prophetの精度が低下しやすい局面ほど生成AI補完の効果が大きい」という知見が、サビキ以外の非連続的な需要を持つ製品でも成立するかを確認する必要がある。

BtoB構造による精度の上限という第二の限界に対しては、可能であれば卸先の販売データを入手し、A社の受注データと組み合わせることで、最終消費者の需要により近い情報を用いた予測を試み、精度の上限がどの程度緩和されうるかを確認する余地がある。

方法論の簡略化という第三の限界に対しては、加法的フォールバックに代えて、Prophetの出力水準に応じて連続的に適用ルールを設計する、より体系的な補正方式を

検討することが課題となる。また、非負制約やゼロ膨張性に対応した時系列モデルを用いれば「ゼロ予測の壁」自体が生じにくくなる可能性がある。本研究が示したゼロ予測の壁は、間欠需要予測（Croston 法（Croston, 1972）・TSB 法（Teunter et al., 2011）等）が扱う問題と本質的に接続するものであり、これら間欠需要に特化した手法を基準モデルとした場合に生成 AI 補完の効果がどう変化するかを検証することは、今後の重要な課題である。基準モデルを XGBoost・LightGBM・LSTM 等に置き換えた場合の効果検証も同様に有益である。

実装上の制約という第四の限界に対しては、API 利用プランの見直しやリクエストの分散処理といった対応策を実際に講じた上で、迅速性がどの程度改善するかを再検証する必要がある。また、A 社の現場担当者へのヒアリングやアンケートを実施し、生成 AI が出力する根拠が実際に納得性の向上や意思決定の迅速化につながるかを定量的に検証すること、および、より大きなサンプルサイズと在庫回転率等の追加指標を用いて欠品機会損失の分析を本格的に実施することも、今後の課題として残される。あわせて、生成 AI の判断根拠が担当者の暗黙知を可視化したものであるという本研究の解釈についても、関係者への開示とフィードバック取得を通じて検証する必要がある。ただし、現時点の精度は実務導入を提案できる水準になく、この段階での開示はかえって手法全体への否定的な評価を招きかねないため、上記の精度改善の取り組みを経て、実務導入の検討が可能な水準に達した段階で改めて実施することとしたい。

生成 AI の判断能力に起因する誤差の悪化という第五の限界に対しては、生成 AI への入力に Prophet の予測値だけでなく、その値がすでに妥当な水準にあると考えられるかどうかの手がかりを加え、Prophet の現状の正確さに応じて補正の要否・程度を調整できるよう、プロンプトや適用ルール自体を精緻化することが今後の課題である。

単一モデル・単発実行という第六の限界に対しては、同一入力に対する複数回の試行によって出力の安定性を定量的に評価するとともに、複数の生成 AI モデルを用いて本研究の知見の再現性を検証することが、今後の課題として残される。

謝辞

本研究の遂行にあたり、データ提供およびインタビューにご協力いただいた A 社に深く感謝申し上げます。

また、本稿の執筆にあたり、終始にわたって懇切丁寧にご指導を賜りました兵庫県立大学大学院社会科学研究科の福田直樹准教授、ならびに、共に学ぶ中でお互いを高めあい、新たな気づきや視点を与えてくれた仲間たちに心より感謝申し上げます。

参考文献

[1] Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 23(3), 289-303.

[2] Taylor, S.J., & Letham, B. (2018). Forecasting at scale. *The American Statistician,* 72(1), 37-45.

[3] Teunter, R. H., Syntetos, A. A., & Babai, M. Z. (2011). Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research,* 214(3), 606-615.

[4] Xie, J. (2016). *Probabilistic Electric Load Forecasting*. Doctoral dissertation, The University of North Carolina at Charlotte.

[5] 伊藤克哉・中川慧(2024)「LLM-Traders：大規模言語モデルを用いた金融時系列予測」人工知能学会全国大会。

[6] 中山 聖也・浦野 昌一(2023)「次元削減を活用した広域気象情報による電力需要予測のための基礎検討」電気学会 電力・エネルギー部門大会論文集。

[7] 永井将太ほか(2025)「販売期間が限られた商品に対するゼロショット需要予測手法」人工知能学会全国大会。

[8] 森山健・井對貴之(2025)「時系列分解と外部データを用いた需要予測方式」人工知能学会全国大会論文集。

[9] 安田大誠・吉野孝・松山浩士(2021)「小売店における商品特性の時系列変動パターンと需要予測精度の関係分析」和歌山大学システム工学研究科。

[10] 山口真一ほか(2024)「データで見る生成 AI の活用と社会的影響」日本経済調査協議会。

[11] 弓岡とら(2023)「Prophet を用いた需要予測在庫管理システムの構築」愛媛大学卒業論文。

[12] 渡邉 信太・小山 賢太郎・青柳 和弘・立川 竜三・小野 由樹子(2020)「電力需要予測業務における機械学習の応用」人工知能学会論文集。

引用ホームページ

[1] SAP News Center「日本ハム、SAP BTP と AI による在庫引当と需要予測で欠品率を大きく改善し、在庫水準を最適化」
https://news.sap.com/japan/2026/04/28045/ (2026 年 4 月 25 日アクセス)

[2] 気象庁「各種データ・資料/過去の気象データ・ダウンロード」
https://www.data.jma.go.jp/risk/obsdl/index.php（2026年6月24日アクセス）
[3] 日本銀行「時系列統計データ検索サイト」
https://www.stat-search.boj.or.jp/（2026年6月29日アクセス）

付録：生成AIへのプロンプト

使用モデル：Claude Sonnet 4.6(claude-sonnet-4-6), max_tokens：1000, temperature等はAPIデフォルト

[System message]

あなたは釣具製造業の需要予測における意思決定支援アシスタントです。モデルベース型時系列予測手法（Prophet、トレンドの柔軟性を高めた設定。年次季節性・祝日効果は明示的には含めていません）が算出した、サビキ仕掛け製品の月次受注数量予測値を、SKUの商品属性および外部環境要因（気温・降水量・為替・輸入物価・祝日）を踏まえて見直してください。
なお、気温・降水量は、このSKUの過去の主要な出荷先都道府県の分布で加重平均した値であり、全国一律の値ではありません。また、Prophetの予測値は0未満にならないよう下限を0としています（0は「ほぼ需要がない」ことを意味し、Prophet自体が0未満の値を算出した月であることを示唆します）。
留意点:
あなたの役割は新たな予測値を独自算出することではなく、Prophetの予測値を「どの方向に」「どの程度」補正すべきかを提案することです。
各月は独立して判断してください。前後の月と同じ補正にする必要はありません。
与えられた情報以外の事後情報（実際の販売実績等）は使わず、与えられた情報のみで判断してください。
根拠が乏しい場合は無理に大きな補正を提案せず、direction="維持"、pctは0に近い値、confidence="低"としてください。
reasonは20文字以内、簡潔な日本語で。
出力は次のJSON形式のみを返してください。説明文・前置き・コードブロック記号は一切付けないこと。

```
{
  "direction": "増加" | "減少" | "維持",
  "pct": 0〜100の整数,
  "confidence": "高" | "中" | "低",
  "reason": "20文字以内の日本語"
}
```

[User message]

【SKU属性】品番|カテゴリー|針サイズ|ハリスサイズ|商品ランク|シリーズ
【対象期間の月次データ】
年月|Prophet予測値|平均気温（主要出荷先加重）|降水量合計（同）|為替レート|輸入物価指数|祝日数|最大連休|大型イベント