

Philippine Higher Education Research Outlook (2025–2035)

A Strategic Forecasting Report for HEI Research Productivity

Metadata	Value
Generated Date	January 2026
Author	IRAP System (Integrated Research Analytics Platform)
Classification	Executive Briefing Document

Abstract

This report provides a comprehensive analysis of research productivity trends across Philippine Higher Education Institutions (HEIs) from 2015 to 2025, with strategic projections extending to 2035. The study employs a **Pandemic-Aware Forecasting Framework** that explicitly accounts for the structural disruption caused by COVID-19 (2020–2022), ensuring that long-term projections are not contaminated by pandemic-induced volatility. Using Holt's Linear Trend method with Simple Moving Average fallback, we project continued growth in research output across most regions, with notable variations in publication quality as measured by Field-Weighted Citation Impact (FWCI).

Methodology

1. The COVID-19 Structural Break

The COVID-19 pandemic (2020–2022) introduced a **structural break** in research productivity data. This period was characterized by:

- Laboratory closures and field research delays
- Transition to remote work affecting collaborative projects
- Publication pipeline disruptions and journal backlogs
- Reallocation of institutional resources to pandemic response

We treat 2020–2022 as a distinct analytical period to isolate pandemic effects from underlying growth trends.

2. Temporal Segmentation Framework

Period	Years	Duration	Characterization
Pre-Pandemic	2015–2019	5 years	Established baseline trends
During Pandemic	2020–2022	3 years	High volatility / Disruption
Post-Pandemic	2023–2025	3 years	"New Normal" recovery
Forecast Phase 1	2026–2030	5 years	Short-term projection
Forecast Phase 2	2031–2035	5 years	Long-term projection

3. Algorithm Selection Logic

The forecasting engine dynamically selects the appropriate model based on data density:

```
For each (School, Metric):  
    Count non-zero observations in training period (2015–2025)  
  
    IF n ≥ 3 non-zero points:  
        → Apply Holt's Linear Trend (captures momentum)  
    ELSE:(2026–2035)  
        → Apply Simple Moving Average (conservative estimate)
```

4. Holt's Linear Trend Method

For institutions with sufficient historical data (≥ 3 non-zero observations), we apply **Holt's Exponential Smoothing** (double exponential smoothing), which decomposes the time series into level and trend components:

Level Equation:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

Trend Equation:

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

Forecast Equation:

$$\hat{Y}_{t+h} = L_t + h \cdot T_t$$

Where:

- Y_t = Observed value at time t
- L_t = Estimated level at time t
- T_t = Estimated trend at time t
- α = Smoothing parameter for level ($0 < \alpha < 1$, optimized via MLE)
- β = Smoothing parameter for trend ($0 < \beta < 1$, optimized via MLE)
- h = Forecast horizon (years ahead)

Implementation: `statsmodels.tsa.holtwinters.Holt`

5. Simple Moving Average (Fallback)

For institutions with sparse data (< 3 non-zero observations), Holt's method risks overfitting or producing unstable forecasts. We apply a **3-period Simple Moving Average**:

$$\hat{Y}_{t+h} = \frac{1}{k} \sum_{i=t-k+1}^t Y_i$$

Where:

- $k = \min(3, n)$ where n is the number of available observations
 - This produces a conservative, flat projection that avoids explosive or negative forecasts
-

6. Post-Processing Constraints

Constraint	Implementation	Rationale
Non-negativity	<code>max(0, forecast)</code>	Publication and citation counts cannot be negative
Discrete rounding	<code>round()</code> for count metrics	Publications and Citations are whole numbers
Continuous FWCI	No rounding	Field-Weighted Citation Impact is a calculated ratio

7. Concrete Examples: Algorithm Selection in Practice

The forecasting system dynamically selects between two methods based on data availability. Below are examples of each approach.

Example A: Holt's Linear Trend (≥ 3 Non-Zero Observations)

Consider **Benguet State University (BSU)** in the Cordillera Administrative Region (CAR). With 11 years of data (2015–2025), BSU has ≥ 3 non-zero observations, qualifying for Holt's method:

Year	Publications	Growth Rate
2023	45	—
2024	52	+15.6%
2025	58	+11.5%

Model Parameters (after fitting):

- $L_{2025} = 58$ (estimated level)
- $T_{2025} = 6.5$ (estimated annual trend)

Forecasts:

- **2026:** $\hat{Y}_{2026} = 58 + 1 \times 6.5 = 64.5 \approx 65$ publications
- **2030:** $\hat{Y}_{2030} = 58 + 5 \times 6.5 = 90.5 \approx 91$ publications

Example B: Simple Moving Average (< 3 Non-Zero Observations)

For institutions with sparse data (fewer than 3 non-zero observations in the training period), the system applies a **Simple Moving Average** to avoid overfitting:

Year	Publications
2023	0
2024	5
2025	7

With only 2 non-zero observations, the SMA fallback is triggered:

$$\hat{Y}_{2026-2035} = \frac{5+7}{2} = 6 \text{ publications/year (constant)}$$

This conservative approach produces a flat projection, preventing explosive or negative forecasts from insufficient trend data.

Disclaimer: These forecasts are statistical projections based on historical trends. They do not account for policy changes, institutional initiatives, or external economic factors. Treat as indicative scenarios.

```
In [1]: # Setup and Data Loading
import sys
import os
import warnings

warnings.filterwarnings('ignore')
sys.path.insert(0, os.path.abspath('..'))

import pandas as pd
import numpy as np

from src.viz_utils import plot_period_geospatial_comparison, PERIOD_ORDER, assign_period
```

```

# Load forecast data
df = pd.read_parquet('../data/processed/forecasts.parquet')

# Add Period and Type columns
df['Period'] = df['Year'].apply(assign_period)
df['Type'] = df['Year'].apply(lambda y: 'History' if y <= 2025 else 'Forecast')

print(f"Dataset loaded: {len(df)} records | {df['School'].nunique()} schools | {df['Region'].nunique()} regions")

```

Dataset loaded: 3,276 records | 52 schools | 17 regions

Summary Statistics by Period

```

In [2]: # Summary by Period with sorted columns (Publication, Citation, FWCI)
summary = df.pivot_table(
    index='Period',
    columns='Metric',
    values='Value',
    aggfunc='sum'
).reindex(PERIOD_ORDER)

# Reorder columns: Publication Quantity, Citation Quantity, Field-Weighted Citation Impact
column_order = ['Publication Quantity', 'Citation Quantity', 'Field-Weighted Citation Impact']
summary = summary[column_order]

styled_summary = (
    summary.style
    .format('{:.0f}')
    .background_gradient(cmap='YlGnBu', axis=0)
    .set_caption('Total Research Output by Period')
)
styled_summary

```

Out[2]:

Total Research Output by Period

Metric	Publication Quantity	Citation Quantity	Field-Weighted Citation Impact
Period			
Pre-Pandemic (2015-2019)	2,278	25,162	125
During Pandemic (2020-2022)	3,649	33,922	128
Post-Pandemic (2023-2025)	6,766	16,078	138
Forecast Phase 1 (2026-2030)	19,652	32,322	376
Forecast Phase 2 (2031-2035)	30,390	37,854	547

Geospatial Analysis

The following animated maps display the regional evolution of research productivity across the five strategic periods. Color intensity and bubble size represent **average annual values**.

Publication Quantity

```
In [3]: fig_pub = plot_period_geospatial_comparison(df, 'Publication Quantity')
fig_pub.show()
```

Citation Quantity

```
In [4]: fig_cit = plot_period_geospatial_comparison(df, 'Citation Quantity')
fig_cit.show()
```

Field-Weighted Citation Impact (FWCI)

```
In [5]: fig_fwci = plot_period_geospatial_comparison(df, 'Field-Weighted Citation Impact')
fig_fwci.show()
```

Complete Forecast Data (2026–2035)

The table below presents forecasted values for all schools, organized in wide format by metric and year.

```
In [6]: # Prepare forecast data for display - Grouped by Year, then Metric
import plotly.graph_objects as go

forecast_df = df[df['Year'] >= 2026].copy()

# Pivot to wide format with Year as first Level, Metric as second
forecast_wide = forecast_df.pivot_table(
    index=['Region Code', 'Region', 'School'],
    columns=['Year', 'Metric'],
    values='Value',
    aggfunc='first'
)

# Flatten MultiIndex columns: 'Year / MetricAbbrev' format
def abbrev_metric(m):
    if 'Publication' in m: return 'Pub'
    elif 'Citation' in m: return 'Cit'
    else: return 'FWCI'

forecast_wide.columns = [f"{year} | {abbrev_metric(metric)}" for year, metric in forecast_wide.columns]
forecast_wide = forecast_wide.reset_index()

# Sort by year, then within each year: Pub, Cit, FWCI
base_cols = ['Region Code', 'Region', 'School']
data_cols = [c for c in forecast_wide.columns if ' | ' in c]
years = sorted(set(int(c.split(' | ')[0]) for c in data_cols))
ordered_cols = []
for year in years:
    for suffix in ['Pub', 'Cit', 'FWCI']:
        col = f"{year} | {suffix}"
        if col in forecast_wide.columns:
            ordered_cols.append(col)

forecast_wide = forecast_wide[base_cols + ordered_cols]
```

```

forecast_wide = forecast_wide.sort_values(['Region Code', 'School']).reset_index(drop=True)

print(f'Forecast table: {len(forecast_wide)} schools x {len(forecast_wide.columns)} columns')
print(f'Years covered: {min(years)} - {max(years)})')
print('Column Key: Pub=Publication, Cit=Citation, FWCI=Field-Weighted Citation Impact')

# Create interactive Plotly table for full visibility
# Use .values.tolist() to handle both Series and DataFrame cases
cell_values = [forecast_wide.iloc[:, i].values.tolist() for i in range(len(forecast_wide.columns))]

fig = go.Figure(data=[go.Table(
    columnwidth=[80, 140, 220] + [70] * len(ordered_cols),
    header=dict(
        values=[f'<b>{col}</b>' for col in forecast_wide.columns],
        fill_color='#1a1a2e',
        font=dict(color='#4ECDC4', size=10),
        align='center',
        height=40,
    ),
    cells=dict(
        values=cell_values,
        fill_color=[[ '#16213e', '#1a1a2e'] * (len(forecast_wide) // 2 + 1)],
        font=dict(color='#e0e0e0', size=10),
        align='center',
        height=25,
    )
)])
fig.update_layout(
    title=dict(
        text='Complete Forecast Data (2026-2035) - Grouped by Year',
        font=dict(color='#4ECDC4', size=16),
        x=0.5
    ),
    paper_bgcolor='#0f0f23',
    margin=dict(l=10, r=10, t=50, b=10),
    height=800,
)
fig.show()

```

```
Forecast table: 52 schools x 33 columns
Years covered: 2026 - 2035
Column Key: Pub=Publication, Cit=Citation, FWCI=Field-Weighted Citation Impact
```

```
In [7]: # Split data by type
historical_df = df[df['Year'] <= 2025].copy()
forecast_df = df[df['Year'] >= 2026].copy()

# Pivot both to wide format for readability
def pivot_to_wide(data, suffix=''):
    """Pivot long-format data to wide format with Year | Metric columns."""
    wide = data.pivot_table(
        index=['Region Code', 'Region', 'School'],
        columns=['Year', 'Metric'],
        values='Value',
        aggfunc='first'
    )

    # Flatten MultiIndex columns
    def abbrev_metric(m):
        if 'Publication' in m: return 'Pub'
        elif 'Citation' in m: return 'Cit'
        else: return 'FWCI'

    wide.columns = [f"{year} | {abbrev_metric(metric)}" for year, metric in wide.columns]
    wide = wide.reset_index()

    # Order columns: base cols + sorted year/metric cols
    base_cols = ['Region Code', 'Region', 'School']
    data_cols = [c for c in wide.columns if ' | ' in c]
    years = sorted(set(int(c.split(' | ')[0]) for c in data_cols))
    ordered_cols = []
    for year in years:
        for s in ['Pub', 'Cit', 'FWCI']:
            col = f"{year} | {s}"
            if col in wide.columns:
                ordered_cols.append(col)

    return wide[base_cols + ordered_cols].sort_values(['Region Code', 'School']).reset_index(drop=True)

historical_wide = pivot_to_wide(historical_df)
```

```

forecast_wide_export = pivot_to_wide(forecast_df)

# Write to Excel with two worksheets
output_path = '../reports/HEI_Research_Report_Data.xlsx'
with pd.ExcelWriter(output_path, engine='openpyxl') as writer:
    historical_wide.to_excel(writer, sheet_name='Historical Figures', index=False)
    forecast_wide_export.to_excel(writer, sheet_name='Forecasted Figures', index=False)

print(f' ✅ Excel report saved to: {output_path}')
print(f' - Historical Figures: {len(historical_wide)} rows x {len(historical_wide.columns)} columns (2015-2025)')
print(f' - Forecasted Figures: {len(forecast_wide_export)} rows x {len(forecast_wide_export.columns)} columns (2026-2035)')

```

- ✓ Excel report saved to: ../reports/HEI_Research_Report_Data.xlsx
 - Historical Figures: 52 rows x 36 columns (2015-2025)
 - Forecasted Figures: 52 rows x 33 columns (2026-2035)

Strategic Outlook

Key Insights

1. **Regional Concentration Risk:** NCR continues to dominate national research output. Strategic investments in regional research centers can diversify the national research portfolio.
2. **Quantity vs. Quality Trade-off:** Regions with high publication growth but low FWCI should prioritize quality assurance mechanisms.
3. **Post-Pandemic Recovery:** The 2023–2025 recovery period shows resilient institutional capacity.
4. **Emerging Research Hubs:** Regions showing steep growth trajectories represent opportunities for targeted capacity-building.

Recommended Actions

Priority	Action Item
High	Establish regional research consortia
High	Implement FWCI-based incentive structures

Priority	Action Item
Medium	Develop pandemic-resilient research continuity plans
Low	Monitor forecast accuracy and recalibrate annually

This report was generated by the IRAP System. Data sourced from CHED and Scopus databases.