

LAPORAN

ANALISIS PREDIKSI PERMINTAAN LISTRIK GLOBAL MENGGUNAKAN REGRESI LINEAR



Oleh:

Ghani Mudzakir 2310817110011

Naufal Elyzar 2310817210019

Noviana Nur Aisyah 2310817120005

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS LUMBUNG MANGKURAT
BANJARMASIN
FEBRUARI, 2026

DAFTAR ISI

DAFTAR ISI.....	1
BAB I: PENDAHULUAN.....	2
1.1 Latar Belakang & Urgensi.....	3
1.2 Deskripsi Dataset.....	3
1.3 Pipeline Pre-processing	5
BAB II: METODOLOGI.....	6
2.1 Keterbaruan Model & Modifikasi Kode	6
2.2 Perancangan Modul Uji Ketahanan (Stress Test)	6
BAB III: HASIL & PEMBAHASAN	6
3.1 Experiment Log & Evaluasi Matriks	7
3.2 Analisis Feature Importance.....	7
BAB IV: STRESS TEST.....	8
4.1 Skenario Stress Test.....	8
4.2 Visualisasi Error	8
BAB V: KESIMPULAN	10
TAUTAN GITHUB.....	10
DAFTAR PUSTAKA	10

BAB I: PENDAHULUAN

1.1 Latar Belakang & Urgensi

Isu pemanasan global (*Global Warming*) telah menjadi krisis eksistensial bagi makhluk hidup di bumi, dengan tahun 2024 tercatat sebagai salah satu periode terpanas dalam sejarah yang melampaui ambang batas kritis yang ditetapkan dalam *Paris Agreement* (Tumwebaze, William, Harold, & Ronica, 2025) (WMO, 2025). Peningkatan suhu bumi ini berkorelasi kuat dengan emisi gas rumah kaca (GRK), termasuk karbon dioksida (CO_2) yang dihasilkan dari aktivitas pembakaran energi fosil dan industri (Sukmawati & Hariyani, 2025). Meningkatnya kebutuhan energi global di berbagai sektor menjadikan prediksi permintaan listrik (*electricity demand*) sebagai salah satu pilar krusial dalam perencanaan pembangunan berkelanjutan (Akbar, 2025). Pertumbuhan ekonomi (GDP) dan populasi suatu negara berbanding lurus dengan konsumsi energinya (Akbar, 2025). Salah satu tantangan dari isu pemanasan global ini adalah penyeimbangan pertumbuhan ekonomi (GDP) dengan keberlanjutan lingkungan (Prastika, 2023). Oleh karena itu, analisis prediktif terhadap pola konsumsi listrik sangat diperlukan untuk menyeimbangkan pertumbuhan ekonomi dengan keberlanjutan lingkungan.

Untuk menganalisis fenomena ini, kami memilih dataset "*Energy & Emissions*" dari *Our World in Data* (OWID). Dataset ini dipilih karena mencakup fitur-fitur prediktor yang sangat relevan, mulai dari indikator makroekonomi (populasi dan GDP), bauran konsumsi energi (fosil, terbarukan, nuklir), hingga tingkat emisi karbon. Rentang data historisnya yang luas dan mencakup berbagai negara memungkinkan algoritma *machine learning* untuk menangkap tren non-linear dan pola konsumsi energi global secara lebih efektif.

Penelitian ini mengimplementasikan model *machine learning* berbasis *ensemble tree* (*Random Forest* dan *XGBoost*) sebagai pembaruan dari model *baseline* (Regresi Linear) untuk memprediksi tingkat permintaan listrik (*electricity demand*). Lebih lanjut, model yang dibangun tidak hanya dievaluasi menggunakan metrik akurasi standar, tetapi juga diuji ketangguhannya melalui rancangan *Stress Test* berbasis deteksi anomali. Pengujian ini bertujuan untuk mengisolasi data ekstrem (*outlier*) akibat lonjakan konsumsi listrik yang tidak wajar di negara-negara tertentu, sehingga dapat membuktikan stabilitas dan performa murni model dalam memberikan gambaran prediktif yang akurat bagi para pembuat kebijakan.

1.2 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini adalah *Data on Energy* yang diterbitkan oleh *Our World in Data* (OWID). Dataset ini diperbarui secara berkala dan mencakup data historis komprehensif mengenai konsumsi energi (*primary energy*), bauran listrik (*electricity mix*), serta metrik relevan lainnya di berbagai negara.

Tabel 1.1 Spesifikasi Dataset

Spesifikasi Dataset	
Sumber	<i>Our World in Data</i> (OWID) – Energy https://github.com/owid/energy-data?tab=readme-ov-file
Jumlah Baris	6.149 baris (setelah proses pembersihan)
Jumlah Fitur	23 Fitur prediktor + 1 target
Target Variabel	<i>electricity_demand</i>

Dataset asli memiliki struktur yang sangat luas. Oleh karena itu, setelah melalui tahap *pre-processing*, penyaringan *data leakage* dan penghapusan *missing value* ekstrem, kami

menetapkan fitur-fitur utama yang paling relevan untuk memprediksi permintaan listrik. Deskripsi rinci mengenai 23 fitur prediktor beserta target variabel tersebut dapat dilihat pada tabel di bawah ini.

Tabel 1.2 Deskripsi *Dataset*

No	Nama Kolom / Fitur	Deskripsi Parameter	Peran
1	electricity_demand	Kebutuhan listrik, diukur dalam terawatt-jam.	Target (Y)
2	year	Tahun observasi.	Fitur (X)
3	population	Populasi berdasarkan negara, tersedia dari 10.000 SM hingga 2100, berdasarkan data dan perkiraan dari berbagai sumber.	Fitur (X)
4	gdp	Total <i>output</i> ekonomi suatu negara atau wilayah per tahun. Data ini disesuaikan dengan inflasi dan perbedaan biaya hidup antar negara.	Fitur (X)
5	biofuel_consumption	Konsumsi energi primer dari bahan bakar nabati (biofuel), diukur dalam terawatt-jam.	Fitur (X)
6	carbon_intensity_elec	Intensitas karbon dari produksi listrik, diukur dalam gram karbon dioksida yang dipancarkan per kilowatt-jam.	Fitur (X)
7	coal_consumption	Konsumsi energi primer dari batubara, diukur dalam terawatt-jam.	Fitur (X)
8	coal_production	Produksi batubara, diukur dalam terawatt-jam.	Fitur (X)
9	energy_per_gdp	Konsumsi energi per unit PDB.	Fitur (X)
10	fossil_fuel_consumption	Konsumsi bahan bakar fosil, diukur dalam terawatt-jam. Ini adalah jumlah energi primer dari batubara, minyak, dan gas.	Fitur (X)
11	gas_consumption	Konsumsi energi primer dari gas, diukur dalam terawatt-jam.	Fitur (X)
12	gas_production	Produksi gas, diukur dalam terawatt-jam.	Fitur (X)
13	greenhouse_gas_emissions	Emisi gas rumah kaca yang dihasilkan dalam pembangkitan listrik, diukur dalam juta ton setara CO ₂ .	Fitur (X)

14	hydro_consumption	Konsumsi energi primer dari tenaga air, diukur dalam terawatt-jam.	Fitur (X)
15	low_carbon_consumption	Konsumsi energi primer dari sumber rendah karbon, diukur dalam terawatt-jam. Ini adalah jumlah energi primer dari energi terbarukan dan tenaga nuklir.	Fitur (X)
16	net_elec_imports	Impor listrik bersih, diukur dalam terawatt-jam.	Fitur (X)
17	nuclear_consumption	Konsumsi energi primer dari tenaga nuklir, diukur dalam terawatt-jam.	Fitur (X)
18	oil_consumption	Konsumsi energi primer dari minyak, diukur dalam terawatt-jam.	Fitur (X)
19	oil_production	Produksi minyak, diukur dalam terawatt-jam.	Fitur (X)
20	other_renewable_consumption	Konsumsi energi primer dari sumber terbarukan lainnya termasuk biofuel, diukur dalam terawatt-jam.	Fitur (X)
21	primary_energy_consumption	Konsumsi energi primer, diukur dalam terawatt-jam.	Fitur (X)
22	renewables_consumption	Konsumsi energi primer dari energi terbarukan, diukur dalam terawatt-jam.	Fitur (X)
23	solar_consumption	Konsumsi energi primer dari tenaga surya, diukur dalam terawatt-jam.	Fitur (X)
24	wind_consumption	Konsumsi energi primer dari tenaga angin, diukur dalam terawatt-jam.	Fitur (X)

1.3 Pipeline Pre-processing

Sebelum data dimasukkan ke dalam model, kami melakukan beberapa tahapan *pre-processing* dengan alur sebagai berikut.

1. *Data Cleaning*

Menghapus baris data yang tidak memiliki nilai pada kolom electricity_demand.

2. Pencegahan *Data Leakage*

Menghapus kolom yang berpotensi membocorkan informasi ke model, seperti semua kolom yang memiliki kata “electricity” atau “generation”.

3. Penyaringan Fitur

Menghapus fitur country dan iso_code, serta menghapus fitur-fitur turunan seperti data per kapita (_per_capita), persentase (_share), dan rasio perubahan (_change).

4. Penanganan *Missing Values*
Mengevaluasi persentase nilai kosong (*null*) pada setiap kolom dan menghapus fitur yang memiliki *missing value* lebih dari 70%.
5. Standarisasi
Menggunakan *library Pipeline* dari Scikit-Learn yang mengintegrasikan SimpleImputer (mengisi sisa nilai kosong dengan nilai median) dan StandardScaler (menyamakan skala seluruh fitur numerik) sebelum proses regresi dilakukan.

BAB II: METODOLOGI

2.1 Keterbaruan Model & Modifikasi Kode

Bentuk keterbaruan dari model *baseline* (Regresi Linear), kami mengimplementasikan dua model regresi lanjutan, yaitu *Random Forest Regressor* dan *XGBoost Regressor*. Alasan utama keterbaruan ini adalah karena Regresi Linear mengasumsikan suatu korelasi yang *linear* atau konstan, padahal data, khususnya data energi ini, dapat memiliki korelasi *non-linear* yang kompleks dan mengandung *outlier*. Keterbaruan ini melakukan penyesuaian *hyperparameter* secara spesifik untuk mencegah *overfitting* dan meningkatkan stabilitas prediksi.

Untuk menambah keterbaruan dari model tersebut, kami tidak hanya menggunakan konfigurasi parameter default, tetapi melakukan beberapa modifikasi eksperimental. Pada model ***Random Forest***, kami melakukan manual *hyperparameter tuning* terhadap parameter *max_depth* (5, 10, 20, dan None), di mana hasil menunjukkan performa stabil dan optimal pada $\text{max_depth} \geq 10$ dengan nilai R^2 mendekati 0.997. Kami juga mengaktifkan parameter *oob_score=True* dan *bootstrap=True* untuk memperoleh *Out-of-Bag (OOB) Score* sebagai validasi internal yang konsisten dengan R^2 pada data uji. Keterbaruan lain pada model ini adalah analisis *feature importance*, di mana fitur *greenhouse_gas_emissions* menunjukkan kontribusi dominan (97%), memberikan insight yang kuat dari sisi interpretabilitas model.

Sementara itu, pada model ***XGBoost Regression***, dilakukan modifikasi *hyperparameter* utama (*n_estimators=500*, *learning_rate=0.05*, *max_depth=6*, *subsample=0.8*, dan *colsample_bytree=0.8*) agar proses *boosting* lebih stabil. XGBoost ini juga dilatih memanfaatkan *eval_set* dan *eval_metric=rmse* untuk memonitor data validasi selama *training*. Secara keseluruhan, eksplorasi kompleksitas model, evaluasi generalisasi (OOB), dan analisis interpretabilitas ini memberikan kedalaman metodologis yang sistematis dan analitis.

2.2 Perancangan Modul Uji Ketahanan (*Stress Test*)

Selain modifikasi pada arsitektur algoritma, keterbaruan lain yang kami usulkan berada pada tahapan evaluasi pasca-pelatihan, yaitu perancangan skenario *Stress Test* secara manual. Modul ini dibangun untuk mendeteksi kerentanan model terhadap anomali data global. Metodologi yang digunakan adalah perhitungan residual (selisih antara nilai prediksi dan nilai aktual), dilanjutkan dengan penerapan batas statistik *Interquartile Range (IQR)* untuk mendeteksi data pencilan (*outlier*). Melalui pendekatan ini, *pipeline* evaluasi tidak hanya mengukur metrik standar secara mentah, tetapi juga mampu mengisolasi *error* yang dihasilkan oleh data ekstrem, sehingga performa murni dari model pada populasi data normal dapat divalidasi dengan lebih objektif.

BAB III: HASIL & PEMBAHASAN

3.1 Experiment Log & Evaluasi Matriks

Berdasarkan serangkaian eksperimen yang telah dilakukan, kami membandingkan performa model *baseline* (Regresi Linear) dengan model regresi lanjutan (*Random Forest* dan *XGBoost*), serta menguji berbagai variasi kedalaman pohon (*max_depth*) pada *Random Forest*. Hasil pengujian dirangkum pada tabel berikut.

Tabel 3.1 *Experiment Log*

model	MAE	RMSE	R2	OOB_Score
DummyRegressor	1052.018	2478.868	-0.00035	
LinearRegression	99.44523	251.5256	0.989701	
RandomForest	26.16352	128.4967	0.997312	
XGBoost	34.80485	158.1999	0.995926	
RandomForest_depth_5	77.82799	231.388	0.991284	0.992811
RandomForest_depth_10	28.834	128.154	0.997326	0.997153
RandomForest_depth_20	26.48079	130.0024	0.997249	0.997284
RandomForest_depth_None	26.16352	128.4967	0.997312	0.997253
XGBoost_Modified	33.33058	156.9354	0.995991	

Dari tabel di atas, terlihat bahwa Regresi Linear sudah memiliki skor yang baik ($R^2 = 0.9897$), namun masih memiliki tingkat kesalahan yang tinggi (RMSE = 251.52). Penerapan model berbasis *ensemble* berhasil menekan tingkat kesalahan tersebut secara signifikan. Model terbaik kami adalah *Random Forest* dengan *max_depth=None*, yang mampu menurunkan RMSE menjadi 128.49 dan mencapai akurasi R^2 tertinggi sebesar 0.9973. Selain itu, nilai *Out-of-Bag (OOB) Score* yang mencapai 0.9972 menunjukkan konsistensi yang berkorelasi dengan skor R^2 , membuktikan bahwa model memiliki kemampuan generalisasi yang sangat baik dan tidak mengalami *overfitting*.

3.2 Analisis Feature Importance

Feature Importance dilakukan untuk memahami interpretasi di balik tingginya akurasi model *Random Forest*. Hasilnya menunjukkan bahwa prediksi permintaan listrik didominasi oleh fitur *greenhouse_gas_emissions* dengan kontribusi sebesar 97.55%. Fitur pendukung lainnya menyumbang persentase yang jauh lebih kecil, seperti *coal_share_elec* (0.30%) dan *nuclear_share_elec* (0.22%). Hal ini mengonfirmasi secara analitis bahwa tingkat emisi gas rumah kaca berbanding lurus dan menjadi prediktor terkuat untuk memetakan lonjakan kebutuhan listrik (*electricity_demand*) suatu negara.

Tabel 3.2 *Feature Importance*

feature	importance
greenhouse_gas_emissions	0.975551
coal_share_elec	0.003021
nuclear_share_elec	0.002259
primary_energy_consumption	0.002054
gas_share_energy	0.001996
other_renewables_share_elec	0.000976
carbon_intensity_elec	0.000932

oil_energy_per_capita	0.000924
nuclear_elec_per_capita	0.000881
hydro_share_elec	0.000727

BAB IV: STRESS TEST

4.1 Skenario Stress Test

Stress test dilakukan untuk menguji robustness model *Random Forest* terhadap data ekstrem atau *outlier*. Meskipun performa awal model sudah sangat tinggi dengan nilai R^2 sebesar 0.997312, diperlukan pengujian tambahan untuk memastikan bahwa performa tersebut tidak hanya dipengaruhi oleh distribusi data tertentu.

Langkah pertama dalam *stress test* adalah menghitung residual pada data uji, yaitu selisih antara nilai aktual dan nilai prediksi. Residual digunakan untuk mengidentifikasi *error model* dan mendeteksi kemungkinan observasi ekstrem. Setelah residual diperoleh, dilakukan deteksi outlier menggunakan metode *Interquartile Range* (IQR), dengan rumus batas bawah dan batas atas sebagai berikut:

- $Lower\ Bound = Q1 - 1.5 \times IQR$
- $Upper\ Bound = Q3 + 1.5 \times IQR$

Observasi yang memiliki residual di luar batas tersebut dikategorikan sebagai *outlier* dan dikeluarkan dari evaluasi lanjutan. Model kemudian diuji kembali pada data yang telah dibersihkan dari *outlier* untuk melihat perubahan performa. Hasil *stress test* menunjukkan:

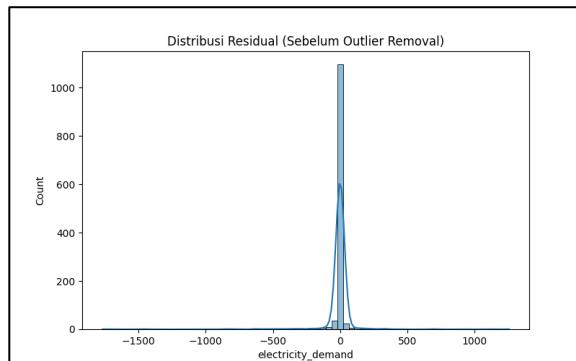
Tabel 4.1 Hasil *Stress Test*

Kondisi	MAE	RMSE	R^2
Sebelum <i>Outlier Removal</i>	26.16352	128.496743	0.997312
Setelah <i>Outlier Removal</i>	0.43604	0.698705	0.999989

Terjadi penurunan error yang sangat signifikan setelah penghapusan *outlier*, dengan MAE turun dari 26.16 menjadi 0.43 dan R^2 meningkat mendekati 1.

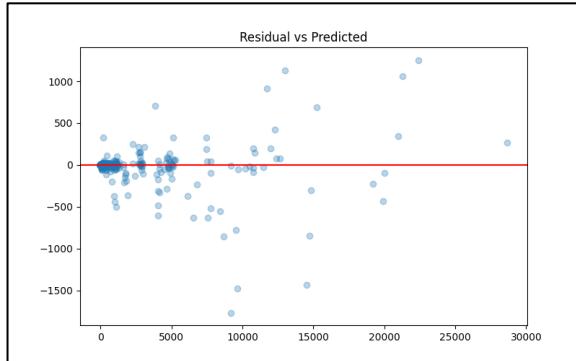
4.2 Visualisasi Error

Visualisasi *error* dilakukan untuk memahami karakteristik residual model serta mengidentifikasi pola kesalahan prediksi sebelum dan setelah penghapusan *outlier*. Analisis ini penting untuk menilai stabilitas dan *robustness* model secara lebih mendalam, tidak hanya berdasarkan nilai metrik evaluasi.



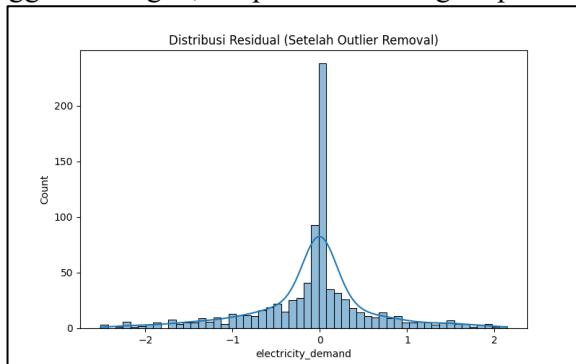
Gambar 4.1 Distribusi Residual (Sebelum *Outlier Remover*)

Pada grafik distribusi residual sebelum *outlier removal* seperti gambar yang ada di atas, terlihat bahwa sebagian besar residual terpusat di sekitar nol, namun terdapat beberapa nilai ekstrem dengan deviasi yang sangat besar. Rentang residual yang lebar menunjukkan adanya observasi dengan *error* signifikan yang berpotensi mempengaruhi nilai RMSE secara drastis, mengingat RMSE sangat sensitif terhadap *error* berskala besar. Meskipun mayoritas prediksi cukup akurat, keberadaan beberapa titik ekstrem menyebabkan distribusi residual memiliki ekor panjang (*long tail*).



Gambar 4.2 Residual vs Predicted

Grafik *residual vs predicted* memperlihatkan bahwa sebagian besar titik tersebar di sekitar garis nol tanpa pola sistematis yang jelas, yang mengindikasikan bahwa model tidak mengalami bias linear tertentu. Namun demikian, terlihat beberapa titik dengan deviasi sangat besar pada nilai prediksi tinggi, yang mengonfirmasi keberadaan outlier pada skala *demand* besar. Hal ini menunjukkan bahwa model relatif stabil pada nilai *demand* rendah hingga menengah, tetapi *error* meningkat pada nilai ekstrem.



Gambar 4.3 Distribusi Residual (Setelah *Outlier Removal*)

Setelah dilakukan penghapusan *outlier* menggunakan metode IQR, distribusi residual menjadi jauh lebih terpusat dan simetris di sekitar nol. Rentang residual menyempit secara signifikan, serta tidak lagi terlihat nilai deviasi ekstrem seperti sebelumnya. Perubahan ini sejalan dengan penurunan drastis nilai MAE dan RMSE pada hasil *stress test*. Visualisasi ini membuktikan bahwa sebagian besar *error* model berasal dari sejumlah kecil observasi ekstrem, sementara pada distribusi data normal model memiliki akurasi yang sangat tinggi dan konsisten.

Secara keseluruhan, visualisasi *error* menunjukkan bahwa model *Random Forest* memiliki kemampuan prediksi yang stabil pada mayoritas data, dengan sensitivitas utama terletak pada observasi ekstrem berskala besar. Dengan demikian, model dapat

dikategorikan robust terhadap distribusi umum data, namun tetap memerlukan perhatian khusus apabila diterapkan pada kondisi dengan nilai *demand* yang sangat tinggi atau tidak lazim.

BAB V: KESIMPULAN

Penelitian ini mengimplementasikan model *machine learning* berbasis *ensemble tree* (*Random Forest* dan *XGBoost*) sebagai pembaruan dari model *baseline* (Regresi Linear) untuk memprediksi tingkat permintaan listrik (*electricity demand*). Model terbaik kami adalah *Random Forest* dengan `max_depth=None` yang mampu menurunkan RMSE menjadi 128.49 dan mencapai akurasi R^2 tertinggi sebesar 0.9973. Hasilnya menunjukkan bahwa prediksi permintaan listrik didominasi oleh fitur `greenhouse_gas_emissions` dengan kontribusi sebesar 97.55%. *Stress test* dilakukan untuk menguji *robustness* model *Random Forest* terhadap data ekstrem atau *outlier*. Terjadi penurunan *error* yang sangat signifikan setelah penghapusan *outlier*, dengan MAE turun dari 26.16 menjadi 0.43 dan R^2 meningkat mendekati 1. Secara keseluruhan, visualisasi *error* menunjukkan bahwa model *Random Forest*ⁱ memiliki kemampuan prediksi yang stabil pada mayoritas data, dengan sensitivitas utama terletak pada observasi ekstrem berskala besar.

TAUTAN GITHUB

Tautan GitHub: <https://github.com/KunyitAlami/Kelompok-9-Pembelajaran-Mesin-1-PREDIKSI-EMISI-CO2-GLOBAL/tree/main>

DAFTAR PUSTAKA

- Akbar, A. A. (2025). *Prediksi Kebutuhan Energi Listrik Kota Semarang Tahun 2025-2030 Berbasis Backpropagation dan Regresi Linear*. Semarang: Program Studi Magister Teknik Elektro Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang.
- Prastika, A. (2023). Hubungan Antara Tingkat Konsumsi Energi Listrik dengan Pertumbuhan Ekonomi di Indonesia. *Jurnal Ilmu Ekonomi (JIE)*, 18-29.
- Sukmawati, M. L., & Hariyani, H. F. (2025). Dampak Investasi Energi Terbarukan terhadap Emisi Karbon di Negara OECD. *Journal of Financial Economics & Investment*, 1-13.
- Tumwebaze, M., William, M., Harold, M., & Ronica, K. B. (2025). Global Warming, a More Catastrophic Climate Change Reality of the 21st Century Than Ever Before: Underlying Factors, Impact and Youth Led Mitigation. *Journal of Geoscience and Environment Protection*, 158-191.
- WMO. (2025, 3 19). *WMO report documents spiralling weather and climate impacts*. Retrieved from World Meteorological Organization: <https://wmo.int/news/media-centre/wmo-report-documents-spiralling-weather-and-climate-impacts>